The Annals of Applied Probability
2020, Vol. 30, No. 2, 706–746
https://doi.org/10.1214/19-AAP1513

TRADING WITH SMALL NONLINEAR PRICE IMPACT

BY THOMAS CAYÉ[1], MARTIN HERDEGEN[2] AND JOHANNES MUHLE-KARBE[3]

[1]*School of Mathematical Sciences, Dublin City University, thomas.caye@dcu.ie*
[2]*Department of Statistics, University of Warwick, M.Herdegen@warwick.ac.uk*
[3]*Department of Mathematics, Imperial College London, j.muhle-karbe@imperial.ac.uk*

We study portfolio choice with small nonlinear price impact on general market dynamics. Using probabilistic techniques and convex duality, we show that the asymptotic optimum can be described explicitly up to the solution of a nonlinear ODE, which identifies the optimal trading speed and the performance loss due to the trading friction. Previous asymptotic results for proportional and quadratic trading costs are obtained as limiting cases. As an illustration, we discuss how nonlinear trading costs affect the pricing and hedging of derivative securities and active portfolio management.

1. Introduction. Classical financial theory is built on the paradigm of *frictionless* markets. By assuming that arbitrary quantities can be traded immediately at the quoted market price, many elegant and far-reaching results can be derived. Real financial markets, however, only supply limited liquidity. Accordingly, execution prices are adversely affected by large trades executed quickly. Optimally scheduling the order flow—to trade off displacement from the optimal frictionless risk-return profile against trading costs—is therefore a crucial concern for large investors such as trend-following hedge funds.

In this paper, we study this problem in a general setting. We consider agents with constant absolute risk aversion,[1] who trade a risky asset with general, not necessarily Markovian, Itô dynamics to maximize their expected utility.

As in the model of Almgren [3], trades incur costs proportional to a power $p \in (1,2)$ of the order flow, corresponding to price impact proportional to the $p-1$-th power of both trade size and execution speed. A price impact elasticity of $p \approx 3/2$ is in line with the "square-root law" advocated by most practitioners (cf., e.g, [6, 47]). The limiting cases $p \to 1$ and $p \to 2$ lead to proportional and quadratic transaction costs—the two frictions that have been the focus of most academic research.[2]

To obtain tractable results in this general setting, we focus on *small* price impact, and perform a sensitivity analysis around the benchmark problem without trading costs. In frictionless diffusion models, optimal trading strategies $\hat{\varphi}_t$ are typically diffusion processes as well, and thereby generate infinite price-impact costs. We show in Theorem 3.3 that, at the leading order and up to stopping close to the terminal time, frictionless target strategies $\hat{\varphi}_t$ of this type are optimally tracked by smoothed strategies φ_t^λ satisfying (pathwise) the following ordinary differential equation (ODE):

$$\dot{\varphi}_t^\lambda = p^{-\frac{1}{p-1}} \left(\frac{\gamma c_t^S}{8\lambda_t}(c_t^{\hat{\varphi}})^2\right)^{\frac{1}{p+2}} \widetilde{g}_p\left(\left(\frac{\gamma c_t^S}{2^{1-p}\lambda_t}(c_t^{\hat{\varphi}})^{-p}\right)^{\frac{1}{p+2}}(\hat{\varphi}_t - \varphi_t^\lambda)\right). \tag{1.1}$$

Received July 2018; revised March 2019.
MSC2010 subject classifications. 91G10, 91G80.
Key words and phrases. Nonlinear price impact, portfolio choice, asymptotics.

[1]Our results formally extend to more general preferences, compare [43]. These arguments could be made rigorous similarly as for proportional transaction costs [1], but we do not pursue this here in order not to drown the new features of the model with nonlinear price impact in (even more) technical estimates arising from random and time-varying risk tolerances.

[2]See, for example, [12, 13, 42, 43, 49, 60] and [5, 8, 12, 13, 26, 27, 31, 50] as well as the references therein for surveys of the large literatures on proportional and quadratic trading costs, respectively.

Here, γ is the agent's risk aversion; $c_t^S = \frac{d\langle S\rangle_t}{dt}$ and $c_t^{\hat{\varphi}} = \frac{d\langle\hat{\varphi}\rangle_t}{dt}$ are the (squared) diffusion coefficients of the risky asset S and the frictionless target strategy $\hat{\varphi}_t$, respectively; p is the elasticity of price impact, and λ_t is the corresponding constant of proportionality describing its magnitude at time t. Finally, the "shape function" $\widetilde{g}_p$ is the rescaled version of the solution g_p of a nonlinear ordinary differential equation; see Section 3.1.

For constant market and preference parameters, (1.1) formally recovers the trading speed that is asymptotically optimal in the Black–Scholes model studied by Guasoni and Weber [32]: a deterministic function of the current deviation from the frictionless optimizer.[3] In our general setting, the asymptotically optimal trading speed remains "myopic", in that it is fully determined by current market and preference parameters, as well as the current displacement from the frictionless target.[4] In particular, the shape function $\widetilde{g}_p$ is universal: it only depends on p, the elasticity of price impact, but not on the other primitives of the model.

In Theorem 3.3, we also calculate the corresponding welfare effects of small price impact. At the leading order, the certainty equivalent loss due to trading costs is

$$c_p \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T \lambda_t^{\frac{2}{p+2}} \left(\frac{\gamma c_t^S (c_t^{\hat{\varphi}})^2}{8}\right)^{\frac{p}{p+2}} dt\right]. \tag{1.2}$$

Here, the constant c_p is obtained from the nonlinear ODE for g_p. The other terms show that price impact has a substantial effect if the market is illiquid (large λ_t), volatile (large c_t^S), or if the frictionless target strategy is difficult to track because it is very active (large $c_t^{\hat{\varphi}}$). If all of these quantities are constant, we formally recover the constant performance loss rate of Guasoni and Weber [32]. When these quantities are time dependent and random, they need to be averaged both across time and states. As for other frictions [42, 43, 50], averaging across states is performed under the frictionless marginal pricing measure $\hat{\mathbb{Q}}$—the small friction is priced like a marginal path-dependent option. The comparative statics of the certainty equivalent loss (1.2) also are consistent with their counterparts for proportional [2, 42], quadratic [50], or fixed transaction costs [7]; the elasticity of price impact p only governs the powers to which the inputs are raised, and determines the universal constant c_p.

To illustrate the wide scope of these results, we discuss two applications where frequent trading is crucial and transaction costs therefore are a prime concern in practice. First, we study the pricing and hedging of derivative securities. Then we turn to the implementation of trend-following investment strategies. Unlike in models with constant investment opportunities, investors cannot "accommodate transaction costs by drastically reducing the frequency and volume of trade" [16] in these settings. Instead, rather frequent trading is necessary and the associated performance losses can be substantial. Formulas (1.1) and (1.2) allow for the first time to study in this context the impact of nonlinear trading costs consistent with the square-root law.

To prove our results, we use a convex duality approach first used in a Mathematical Finance context by Henderson [33] for the indifference pricing of small unhedgeable claims.[5] More

[3] Since we use an absolute parametrization (for risk aversion, returns, etc.), it is difficult to make this connection to the relative quantities of [32] precise; compare [29] for more details.

[4] In contrast, the *exact* optimal optimal strategies have been found to "aim in front of the frictionless target" in models with quadratic costs and preferences, where they can be computed in closed form [8, 26, 27]. The "aim portfolio" in these models is a weighted average of the expected future values of the frictionless target; as the transaction cost tends to zero, it converges to the current value of the latter, leading to a myopic optimal trading speed. Guasoni and Weber [32] study numerically the difference between the performances of the exact and asymptotic optimal policies in a concrete model and find it to be negligible for empirically relevant values of the price impact parameter.

[5] Extensions of these results have been developed by [45]; similar arguments have also been used for the perturbation analysis of small variations of market prices of risk [46, 51] or cumulative random endowments [35].

specifically, we obtain a lower bound for the value expansion (1.2) by analyzing a specific family of trading strategies. An upper bound can in turn be determined using convex duality. For proportional transaction costs, starting with the seminal work of Cvitanić and Karatzas [17], the corresponding duality has been studied intensely, and is used by [1, 2, 34] to obtain tight upper bounds for (1.2). In the present context of superlinear trading costs, abstract duality results have only very recently been developed by Guasoni and Rásonyi [30]. The present study is the first application of these results to a concrete portfolio choice problem. Here, the key challenge is to come up with a concrete candidate for the asymptotic dual minimizer. The first step is to use the first-order condition of [30] to derive a "naive" dual minimizer from our candidate for the asymptotic primal maximizer. The second step is to suitably modify this naive dual candidate in order to control the remainder terms appearing in the asymptotic verification. For the case of proportional transaction costs [1, 2, 34], this can be done by stopping the "naive" dual candidate in an appropriate manner. With superlinear context, this is not sufficient and another subtle modification is necessary.

The computation of the primal and dual bounds proceeds in several steps. We first perform a second-order expansion of the primal and dual goal functionals, thereby reducing them to linear-quadratic functionals.[6] After renormalizing time and space appropriately, we then show that "locally," that is, on each small time interval, this simplified criterion converges to an ergodic mean-variance functional of a controlled diffusion process. In the present context, this controlled process is an Ornstein–Uhlenbeck-type process, with constant volatility but nonlinear mean-reversion speed governed by the function $\widetilde{g}_p$. This makes the analysis much more challenging than in the limiting cases of quadratic trading costs, where the limiting process is a standard Ornstein–Uhlenbeck process, or proportional costs, where the limiting process is a doubly reflected Brownian motion. For the present study, a number of delicate probabilistic estimates need to be developed from scratch to establish convergence to the limiting problem. This is done in the companion paper of the present study [15].

The connection between asymptotics of utility maximization problems with small transaction costs and ergodic control problems with "frozen coefficients" has first been established using PDE techniques by Soner and Touzi [60]. For proportional costs, non-Markovian extensions of these results have been obtained formally by [42, 43] and proved rigorously by [1, 2, 34]. A closely related strand of research studies "pathwise" criteria, where the goal is to trade off the error of tracking an exogenous target strategy against the trading costs incurred by the hedge. Building on work of [24, 25, 28], Cai, Rosenbaum and Tankov [12, 13] study such problems for quite general specifications of tracking errors and trading costs. Using weak-convergence techniques, they derive tight bounds for a number of examples, including proportional and quadratic costs.

This article is organized as follows. In Section 2, we introduce the model. Section 3 presents our main results and the assumptions needed to carry out the rigorous asymptotic analysis. It also contains a discussion of the implications of our results for option pricing and hedging, as well as for active portfolio management. The proof of the main result can be found in Sections 4 and 5: a primal lower bound is derived in Section 4, and the corresponding dual upper bound is constructed in Section 5. For better readability, some auxiliary estimates used in various proofs are delegated to the Appendices.

Notation. For an Itô process X, we write $\mu_t^{X,\mathbb{P}}$ and $\mu_t^{X,\hat{\mathbb{Q}}}$ for the drift rate under the physical measure $\mathbb{P}$ and the marginal pricing measure $\hat{\mathbb{Q}}$ from Section 2.2, respectively. The infinitesimal covariation between two Itô processes X and Y is denoted by $c_t^{X,Y} = \frac{d\langle X,Y\rangle_t}{dt}$,

[6] Such simpler performance criteria are directly used in a number of papers, for example, [5, 8, 12, 13, 26, 27, 56]. The same simplification for more general utilities also obtains for proportional costs [37, 43, 55, 60].

and we write c^X for $c^{X,X}$. We denote the set of all predictable and X-integrable processes H satisfying $\mathbb{E}_{\hat{\mathbb{Q}}}[\int_0^T H_t^2 \, d\langle X\rangle_t] < \infty$ by $L^2_{\hat{\mathbb{Q}}}(X)$. For a continuous process X, we write $X_t^* = \max\{|X_s| : s \in [0,t]\}$ for the corresponding running maximum. The stochastic exponential of an Itô process X is denoted by $\mathcal{E}(X) = \exp(X - \frac{1}{2}\langle X\rangle)$. We write $[\![\sigma,\tau]\!]$ for the stochastic interval between two stopping times σ and τ. Finally, we use the Landau notation in the following way: the symbols $O(\cdot)$ and $o(\cdot)$ refer to pointwise limits, where the asymptotic parameter λ tends to zero.

2. Model.

2.1. *Financial market.* Fix a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ satisfying the usual conditions and supporting at least one standard $\mathbb{P}$-Brownian motion $W^{S,\mathbb{P}}$. We consider a financial market with two assets.[7] The first one is safe, with price normalized to one. The second is risky, with general—not necessarily Markovian—Itô dynamics given by

$$dS_t = \mu_t^{S,\mathbb{P}} \, dt + \sqrt{c_t^S} \, dW_t^{S,\mathbb{P}}, \quad S_0 = s. \tag{2.1}$$

Here, the drift rate $\mu^{S,\mathbb{P}}$ and the (squared) diffusion coefficient $c^S > 0$ are adapted processes for which the stochastic integral (2.1) is well defined on $[0,T]$, and $s \in \mathbb{R}$.

2.2. *Frictionless portfolio choice.* In the above market, we study the portfolio choice problem of an agent with constant absolute risk aversion $\gamma > 0$. To wit, starting from an initial endowment $x \in \mathbb{R}$, the agent's goal is to choose a predictable trading strategy φ to maximize $J^0(\varphi)$, the expected exponential utility from terminal wealth:[8]

$$J^0(\varphi) := \mathbb{E}_{\mathbb{P}}\left[U\left(x + \int_0^T \varphi_t \, dS_t\right)\right] \to \max! \quad \text{where } U(x) = -\exp(-\gamma x). \tag{2.2}$$

To ensure well-posedness of the maximization problem (2.2), we impose the following standard no-arbitrage condition [20, 39, 57, 58].

ASSUMPTION 1. There exists an equivalent local martingale measure $\mathbb{Q}$ for S, which has finite relative entropy with respect to $\mathbb{P}$: $H(\mathbb{Q}|\mathbb{P}) := \mathbb{E}_{\mathbb{P}}[\frac{d\mathbb{Q}}{d\mathbb{P}} \log(\frac{d\mathbb{Q}}{d\mathbb{P}})] < \infty$.

By [23], Theorems 2.1, 2.2 and Remark 2.1, Assumption 1 implies that there exists a local martingale measure $\hat{\mathbb{Q}}$ equivalent to $\mathbb{P}$ that is the unique solution of the dual problem of minimizing the relative entropy with respect to $\mathbb{P}$ among the absolutely continuous local martingale measures.

Assumption 1 also ensures the existence of an optimizer $\hat{\varphi}$ for (2.2) over all S-integrable processes φ whose gains processes $\int_0^\cdot \varphi_t \, dS_t$ are $\hat{\mathbb{Q}}$-martingales [20], Theorem 1. Henceforth, we therefore focus on such *admissible* trading strategies and denote by Φ^0 the set of these. The primal maximizer $\hat{\varphi}$ is linked to the "minimal-entropy martingale measure" $\hat{\mathbb{Q}}$ by the first-order condition of convex duality [58], equation (12):

$$U'\left(x + \int_0^T \hat{\varphi}_t \, dS_t\right) = \hat{y}\frac{d\hat{\mathbb{Q}}}{d\mathbb{P}} \quad \text{for a constant } \hat{y} > 0. \tag{2.3}$$

[7] The probabilistic approach in this paper crucially exploits that the invariant distributions of *scalar* diffusion processes are readily available. In a multidimensional Markovian setting, a small-cost expansion of the value function is studied using stability results for viscosity solutions of the corresponding dynamic programming equations in [10].

[8] As is well known [20], random endowments such as labour income or option positions can be readily absorbed into a change of measure.

2.3. *Portfolio choice with superlinear transaction costs.* As in [4, 6, 30, 32], we now assume that trades incur superlinear costs levied on the trading rate $\dot{\varphi}_t = \frac{d}{dt}\varphi_t$, that is, trading costs increase with both trade size and speed.

More specifically, execution prices are shifted proportionally to a power $p-1 \in (0,1)$ of $\dot{\varphi}_t$, so that the trading costs accumulate at rate $\lambda_t|\dot{\varphi}_t|^p$. Here, the proportionality factor is of the form

$$\lambda_t = \lambda \Lambda_t,$$

for some small parameter $\lambda > 0$ that measures the magnitude of the trading costs, and a positive, continuous Itô process Λ that describes their dynamics. The constant p is the "elasticity of price impact"; proportional costs correspond to the limiting case $p \to 1$, linear price impact (quadratic costs) to $p \to 2$. Empirical studies typically estimate values $p \approx 3/2$; compare [6, 47].

With trading costs, we need to specify how the agent's initial endowment x is allocated between her safe and risky accounts. For simplicity, we assume that the initial risky allocation equals the frictionless optimum $\hat{\varphi}_0$, so that $x_0 = x - \hat{\varphi}_0 S_0$ is the corresponding initial safe position.

Likewise, different terminal conditions are possible, compare [8]. Here, as in [30], we impose that the risky position is eventually liquidated for consumption ($\varphi_T = 0$). The set of all absolutely continuous trading strategies $d\varphi_t = \dot{\varphi}_t\, dt$ that fulfil the above requirements and satisfy $\mathbb{E}_{\hat{\mathbb{Q}}}[(\int_0^T \varphi_t^2 c_t^S\, dt)^{1+a}]$, for some $a > 0$, is denoted by Φ^λ. The frictional wealth process corresponding to such an *admissible* strategy $\varphi \in \Phi^\lambda$ is

$$X_t^\varphi = x + \int_0^t \varphi_s\, dS_s - \int_0^t \lambda_s |\dot{\varphi}_s|^p\, ds, \quad t \in [0,T].$$

Accordingly, in analogy to the frictionless case (2.2), the agent chooses $\varphi \in \Phi^\lambda$ to maximize

$$J^\lambda(\varphi) := \mathbb{E}_{\mathbb{P}}\left[U\left(x + \int_0^T \varphi_t\, dS_t - \int_0^T \lambda_t |\dot{\varphi}_t|^p\, dt\right)\right] \to \max! \tag{2.4}$$

3. Main results. The frictional portfolio choice problem (2.4) is intractable even in the simplest concrete models. We therefore study the asymptotic regime where the magnitude λ of the trading costs tends to zero. Results of this kind have recently been obtained by Guasoni and Weber [32] for a long-term portfolio-choice problem in a Black–Scholes model with scale-invariant price impact. Here, we perform this sensitivity analysis in a general setting. This reveals the general underlying structure of the problem and identifies the relevant statistics that measure the susceptibility of trading strategies to small trading costs. Moreover, as discussed in Section 3.4 below, this allows to treat as special cases the trading problems where transaction costs are most relevant in practice: pricing and hedging of derivative securities and active portfolio management.

3.1. *Asymptotically optimal strategies.* In this section, we define a family of trading strategies $(\varphi_t^\lambda)_{t\in[0,T]}$ that is asymptotically optimal for (2.4) in the limit for small transaction costs λ; see Theorem 3.3. Since the rigorous definition is rather subtle, we proceed in four steps.

A nonlinear ODE. A first ingredient for the definition of φ^λ is the solution to a nonlinear ODE, that also plays a central role in the work of Guasoni and Weber [32], Lemmas 19 and 21.

LEMMA 3.1. *Let $p \in (1,2]$. Then there exists a unique positive constant c_p such that the ordinary differential equation*

$$g_p'(z) = (p-1)p^{-\frac{p}{p-1}}|g_p(z)|^{\frac{p}{p-1}} - z^2 + c_p \tag{3.1}$$

has a solution on $\mathbb{R}$ which satisfies the following growth conditions:

$$\begin{aligned} &\lim_{z\to-\infty} \frac{g_p(z)}{|z|^{\frac{2(p-1)}{p}}} = -p(p-1)^{-\frac{p-1}{p}} \quad \textit{and} \\ &\lim_{z\to+\infty} \frac{g_p(z)}{|z|^{\frac{2(p-1)}{p}}} = p(p-1)^{-\frac{p-1}{p}}. \end{aligned} \tag{3.2}$$

This solution is unique, and is an odd, increasing function.

REMARK 3.2. The growth conditions for (3.1) at $\pm\infty$ are somewhat ad hoc. An alternative characterization is that the constant c_p is the smallest value for which the ODE (3.1) has a solution on the whole real line that is positive on $(0,\infty)$ and negative on $(-\infty, 0)$; see [14], Lemma 2.2.2.

Up to rescaling, the function g_p parametrizes the asymptotically optimal trading speed as a function of the displacement of the frictionless target from the actual position; cf. Theorem 3.3. Accordingly, positivity on $(0,\infty)$ and negativity on $(-\infty,0)$ translate to the natural property that one always trades toward the frictionless optimum. The constant c_p will be seen to describe the size of the corresponding utility loss due to trading costs. Since this needs to be minimized at the optimum, it is natural that the smallest possible choice is the correct one.

Auxiliary SDEs. We next define a family of SDEs that (up to stopping close to the terminal time) describe the asymptotically optimal displacement from the frictionless strategy $\hat{\varphi}$. To simplify notation, we pass to the following rescaled version of the function g_p from (3.1):

$$\widetilde{g}_p(x) = \operatorname{sgn}(x)|g_p(x)|^{\frac{1}{p-1}}, \quad x \in \mathbb{R}. \tag{3.3}$$

Moreover, we impose the following Itô process assumptions.

ASSUMPTION 2. The processes S, $\hat{\varphi}$, Λ, c^S and $c^{\hat{\varphi}}$ are Itô processes with locally bounded drift rates (under $\hat{\mathbb{Q}}$) and locally bounded (squared) diffusion coefficients. We also require that $c^{\hat{\varphi}} > 0$.[9]

Under Assumption 2, it follows from [15], Proposition 1.1, that the SDE

$$\begin{aligned} d\overline{\Delta\varphi}_t^\lambda = {}& \left(\mu_t^{\hat{\varphi},\hat{\mathbb{Q}}} - \frac{1}{2}p^{-\frac{1}{p-1}}\lambda^{-\frac{1}{p+2}}c_t^{\hat{\varphi}}m_t\widetilde{g}_p\left(\lambda^{-\frac{1}{p+2}}m_t\overline{\Delta\varphi}_t^\lambda\right)\right)dt \\ &+ \sqrt{c_t^{\hat{\varphi}}}\,dW_t^{\hat{\varphi},\hat{\mathbb{Q}}}, \quad \overline{\Delta\varphi}_0^\lambda = 0, \end{aligned} \tag{3.4}$$

[9]Note that Assumption 2 implies that the process m defined in (3.7) and the process A defined in (5.4) are again Itô processes with locally bounded drift rates (under $\hat{\mathbb{Q}}$) and locally bounded (squared) diffusion coefficients. To see this, use that if X, Y are Itô processes with locally bounded drift rates (under $\hat{\mathbb{Q}}$) and locally bounded (squared) diffusion coefficients and f is a C^2 function (with appropriate domain), then Itô's formula, the Kunita–Watanabe inequality, and continuity of X, Y imply that $f(X,Y)$ is again an Itô process with locally bounded drift rates (under $\hat{\mathbb{Q}}$) and locally bounded (squared) diffusion coefficients.

has a unique strong solution on $[0, T]$ for all $\lambda > 0$, where $W^{\hat{\varphi},\hat{\mathbb{Q}}}$ is a Brownian motion under $\hat{\mathbb{Q}}$.[10]

We can now define the trading strategy $(\overline{\varphi}_t^\lambda)_{t\in[0,T]}$ by

$$\overline{\varphi}_t^\lambda := \hat{\varphi} - \overline{\Delta\varphi}^\lambda. \tag{3.5}$$

Note that $\overline{\varphi}^\lambda$ satisfies the pathwise ODE

$$\dot{\overline{\varphi}}_t^\lambda = \frac{1}{2} p^{-\frac{1}{p-1}} \lambda^{-\frac{1}{p+2}} c_t^{\hat{\varphi}} m_t \widetilde{g}_p\big(\lambda^{-\frac{1}{p+2}} m_t (\hat{\varphi}_t - \overline{\varphi}_t^\lambda)\big), \quad \varphi_0^\lambda = \hat{\varphi}_0, \tag{3.6}$$

where

$$m_t = \left(\frac{2^{p-1}\gamma c_t^S}{\Lambda_t (c_t^{\hat{\varphi}})^p}\right)^{\frac{1}{p+2}}. \tag{3.7}$$

Stopping times. The strategies $\overline{\varphi}^\lambda$ from (3.5) are essentially asymptotically optimal for (2.4). However, to obtain rigorous asymptotic results and to ensure admissibility, we need to slightly modify $\overline{\varphi}^\lambda$ by appropriate stopping close to the terminal time.

On the one hand, like for models with proportional and fixed transaction costs [2, 22], excessive deviations from the frictionless target need to be avoided. Therefore, liquidation is initiated immediately if the deviation from the frictionless target becomes too large or efficient tracking becomes impossible because the process m becomes too small.[11] The probability of these events is negligible for small λ (see Proposition C.5), but the stopping is crucial to control the remainder terms. On the other hand, we need to ensure here that the risky position is indeed liquidated until maturity T.[12]

To make all this precise, define the time at which the liquidation of the risky position is initiated at the latest by

$$T^\lambda = T - \lambda^\eta \quad \text{where } \eta \in \frac{1}{p+2}\left(2, \frac{p}{p-1} \wedge \frac{8}{3}\right), \tag{3.8}$$

choose constants such that[13]

$$\begin{aligned} &\kappa_1 \in \left(\frac{2}{3}\frac{1}{p+2}, \frac{1}{p+2}\right), \qquad \kappa_2 \in \left(0, \frac{1}{6}\left(\frac{1}{p+2} - \kappa_1\right)\right), \\ &\kappa_3 \in \left(\frac{2}{3}\frac{1}{p+2}, \frac{6-2p}{3}\frac{1}{p+2}\right), \\ &\kappa_4 \in \left(0, \left(\frac{1}{p+2} - \frac{p-1}{p}\eta\right) \wedge \frac{1}{3}\frac{1}{p+2}\right), \end{aligned} \tag{3.9}$$

and define the stopping time

$$\begin{aligned} \tau^{\Delta\varphi} = \inf\Big\{ &t \in [0, T^\lambda] : \big|\overline{\Delta\varphi}_t^\lambda\big| > \lambda^{\kappa_1} \text{ or } \frac{1}{2} p^{-\frac{1}{p-1}} m_t < \lambda^{\kappa_2} \text{ or } m_t > \lambda^{-\kappa_2} \\ &\text{or } \lambda\Lambda_t \big|\dot{\overline{\varphi}}_t^\lambda\big|^p > \lambda^{\kappa_3} \text{ or } |\hat{\varphi}_t| \geq \lambda^{-\kappa_4} \Big\} \wedge T^\lambda. \end{aligned} \tag{3.10}$$

---

[10]In the notation of [15], take $\varepsilon = \lambda^{\frac{1}{p+2}}$, $b = \mu^{\hat{\varphi}}$, $M = m$, $L = \frac{1}{2} p^{-\frac{1}{p-1}} m = \frac{1}{2} p^{-\frac{1}{p-1}} M$, $c = c^{\hat{\varphi}}$.

[11]This happens if the trading cost becomes too large, or the target too volatile relative to the risky asset.

[12]For proportional or fixed costs, this can be done by a single bulk trade at T, without affecting the asymptotic results at the leading order.

[13]Note that this is always possible since $p \in (1, 2)$.

Candidate strategies. We can finally define our candidate strategies φ^λ for $\lambda > 0$ by

$$\varphi_t^\lambda = \overline{\varphi}_t^\lambda \mathbb{1}_{\{t \le \tau^{\Delta\varphi}\}} + \big(1 - \lambda^{-\eta}(t - \tau^{\Delta\varphi})\big)\overline{\varphi}^\lambda_{\tau^{\Delta\varphi}} \mathbb{1}_{\{\tau^{\Delta\varphi} < t \le \tau^{\Delta\varphi} + \lambda^\eta\}}, \quad t \in [0, T]. \tag{3.11}$$

The displacement of the frictionless target from this process is $\Delta\varphi^\lambda = \hat{\varphi} - \varphi^\lambda$. On $[\![0, \tau^{\Delta\varphi}]\!]$, the difference $\Delta\varphi^\lambda$ then coincides with $\overline{\Delta\varphi}^\lambda$ by construction. Moreover, (3.11) guarantees that:

(i) On $[0, \tau^{\Delta\varphi}]$, the trading rate $\dot{\varphi}_t^\lambda$ is determined by the ODE (3.6);
(ii) On $(\tau^{\Delta\varphi}, \tau^{\Delta\varphi} + \lambda^\eta]$ the risky position is liquidated at the constant rate $\dot{\varphi}_t^\lambda = -\lambda^\eta \varphi^\lambda_{\tau^{\Delta\varphi}}$;
(iii) On $(\tau^{\Delta\varphi} + \lambda^\eta, T]$, there are no more trades ($\dot{\varphi}_t^\lambda = 0$) and the agent's position is $\varphi_t^\lambda = 0$.

3.2. *Main result.* For the validity of our main result, the primitives of the model need to be integrable enough. To succinctly formulate the precise conditions, we introduce, for $\varepsilon > 0$, the following set:

$$\begin{aligned}
\mathcal{X}^\varepsilon = \bigg\{&\exp\bigg(8\int_0^T \frac{(\mu_t^{\hat{\varphi},\hat{\mathbb{Q}}})^2}{c_t^{\hat{\varphi}}}\,dt\bigg), \big(\big(m(1 \wedge c^{\hat{\varphi}})\big)^{-16\frac{2+p}{2-p}(1+\varepsilon)}\big)_T^*, \big(c_T^{\hat{\varphi}*}\big)^{16p(1+\varepsilon)}, \\
&\big(\big((\hat{\varphi}_T^*)^2 + 1\big)c_T^{S*}\big)^3, \exp(\varepsilon \Lambda_T^*), \big(c_T^{\hat{\varphi}*}\big(1 + (m_T^*)^3\big)\big)^{\frac{4(1+2\varepsilon)(1+\varepsilon)p}{2-2p+(p+2)(2p\kappa_1-\kappa_3)}}, \\
&\big(m^{\frac{4(1+2\varepsilon)}{(p+2)\kappa_2}}\big)_T^*, \big(m^{-\frac{4(1+2\varepsilon)}{(p+2)\kappa_2}}\big)_T^*, \big(\hat{\varphi}_T^*\big)^{\frac{4(1+2\varepsilon)}{(p+2)\kappa_4}}, \big(\Lambda^{-\frac{1+\varepsilon}{p-1}}\big)_T^*, \\
&\exp\bigg(32\gamma^2 \int_0^T (\hat{\varphi}_t^*)^2 c_t^S\,dt\bigg), \exp\big(\varepsilon\big((c^S)^{-1}\big)_T^*\big)\bigg\}.
\end{aligned}$$

We then impose the following integrability assumptions that are for example satisfied if Λ, c^S, $\hat{\varphi}$, $\mu^{\hat{\varphi},\hat{\mathbb{Q}}}$ and $c^{\hat{\varphi}}$ are bounded and Λ, c^S, $c^{\hat{\varphi}}$ are bounded away from 0.

ASSUMPTION 3. For some $\varepsilon > 0$, we have $\mathcal{X}^\varepsilon \subset L^1(\hat{\mathbb{Q}})$.

We can now formulate our main result. For better readability, its long and technical proof is deferred to Sections 4 and 5.

THEOREM 3.3. *Suppose the no-arbitrage Assumption 1 holds, and the primitives of the model satisfy the Itô process and integrability Assumptions 2 and 3. Then the strategy φ^λ from (3.11) is admissible and* asymptotically optimal *for the frictional utility maximization problem (2.4), in that*

$$\begin{aligned}
\sup_{\varphi \in \Phi^\lambda} J^\lambda(\varphi) &= J^\lambda(\varphi^\lambda) + o\big(\lambda^{\frac{2}{p+2}}\big) \\
&= J^0(\hat{\varphi}) - \hat{y}c_p \mathbb{E}_{\hat{\mathbb{Q}}}\bigg[\int_0^T \lambda_t^{\frac{2}{p+2}} \bigg(\frac{\gamma c_t^S (c_t^{\hat{\varphi}})^2}{8}\bigg)^{\frac{p}{p+2}} dt\bigg] + o\big(\lambda^{\frac{2}{p+2}}\big).
\end{aligned} \tag{3.12}$$

3.3. *Interpretation and comparative statics.* Let us now discuss the interpretation of Theorem 3.3 and look at the limiting cases $p \uparrow 2$ (linear price impact) and $p \downarrow 1$ (proportional transaction cost).

Asymptotically optimal strategies. First, we discuss the comparative statics of the optimal trading strategies.[14]

Similar to models with linear price impact [5, 27, 31, 50], the family of strategies $(\varphi^\lambda)_{\lambda>0}$ track the frictionless target portfolio $\hat{\varphi}$. Their fine structure in turn depends on the degree of activity exhibited by $\hat{\varphi}$.[15]

The trading speed (3.6) is determined by the diffusion coefficients c_t^S and $c_t^{\hat{\varphi}}$ of the risky asset and the frictionless optimizer, the current trading cost λ_t, the risk aversion γ, and the elasticity of the price impact p. In addition, the shape function $\widetilde{g}_p$ determines how the deviation $\Delta\varphi_t^\lambda = \hat{\varphi}_t - \varphi_t^\lambda$ of the frictionless optimizer from the actual position is incorporated.

For $p \uparrow 2$, the shape function $\widetilde{g}_p$ converges to $\widetilde{g}_2(x) = 2x$. The optimal trading speed then simplifies to $\sqrt{\gamma c_t^S/2\lambda_t}$ times the deviation of the frictional portfolio from the target [50], which no longer depends on the variation $c_t^{\hat{\varphi}}$ of the target strategy.

For $p \downarrow 1$, $\widetilde{g}_p$ converges to 0 on $[0, (3/2)^{2/3})$ and to infinity on $((3/2)^{2/3}, \infty]$; cf. [32], Lemmas 8 and 9. As a consequence, the associated trading speed (1.1) explodes once the deviation from the frictionless target exceeds

$$\hat{\varphi}_t - \varphi_t^\lambda = \pm\left(\frac{3}{2\gamma}\frac{c_t^{\hat{\varphi}}}{c_t^S}\lambda_t\right)^{1/3},$$

and it converges to zero between these boundaries. This corresponds to the instantaneous reflection off these trading boundaries that is asymptotically optimal for small *proportional* transaction costs [12, 42, 43, 49, 60].

To understand the comparative statics of the trading rate for general $p \in (1, 2)$, recall that the function $\widetilde{g}_p$ is increasing. Whence, the trading rate (3.6) remains increasing in the ratio of price volatility times risk aversion divided by the current price impact, like for quadratic trading costs [50]. The dependence on the volatility of the frictionless target is more complex. To wit, if the displacement of the frictional position is close to zero, then the ODE (3.1) shows that the function $\widetilde{g}_p$ is proportional to $x \mapsto x^{\frac{1}{p-1}}$. Hence, the trading rate (3.6) is approximately proportional to $c_t^{\hat{\varphi}}$ raised to the power $\frac{2}{p+2} - \frac{p}{(p-1)(p+2)} < 0$. Thus, a large target volatility discourages the agent from trading when she has almost the optimal number of risky shares. The reason is that a price impact elasticity of $p \in (1, 2)$ leads to higher than quadratic trading costs for small trades. Whence, high tracking speeds are reduced near the frictionless optimum. On the other hand, when the displacement is large, the function $\widetilde{g}_p$ scales like $x \mapsto x^{\frac{2}{p}}$ (cf. (3.2)), so that—as in the case of quadratic costs—the trading rate (3.6) no longer depends on $c_t^{\hat{\varphi}}$. The intuition for this is that if the displacement is very large, the volatility of the target becomes insignificant relative to the displacement from the frictionless optimum.

Leading order loss of utility. Next, we discuss the welfare effects of small nonlinear price impact. The first term on the right-hand side of (3.12) is the performance of the frictionless optimizer. Accordingly, the second term corresponds to the minimal leading-order loss that can be achieved by applying the policy from Section 3.1.

This minimal performance loss is of order $O(\lambda^{\frac{2}{p+2}})$ for small trading costs λ. In the limiting cases $p \downarrow 1$ (proportional costs) and $p \uparrow 2$ (quadratic costs), the well-known orders $O(\lambda^{1/3})$ (cf. [37, 55]) and $O(\lambda^{1/2})$ (cf. [31, 50]) from the literature obtain.

[14] Note that since the stopping time τ^φ from (3.10) is very close to the terminal time T, φ^λ can be identified with $\overline{\varphi}^\lambda$ from (3.5) in the following discussion without loss of generality.

[15] For example, if the target strategy is smoother than Brownian motion, then it can be tracked much more closely and with substantially smaller trading costs; compare [56].

The factor multiplying this power of the trading cost has three components: the frictionless Lagrange multiplier $\hat{y}$, the constant c_p from Lemma 3.1, and an average of the model parameters with respect to time and randomness. In view of, for example, [58], Theorem 1, $\hat{y}$ is the derivative of the frictionless performance with respect to the initial endowment. Whence, by Taylor's theorem, the other terms in the leading-order loss can be interpreted as a "certainty-equivalent loss" as in [2, 43, 50]. This means that they correspond to the amount of initial endowment the agent would give up in order to trade the risky asset without transaction costs.

The first ingredient for this "cash equivalent of the small friction" is the constant c_p, which is universal in that it only depends on the elasticity of price impact p but none of the other model parameters. Its limiting values for $p \downarrow 1$ and $p \uparrow 2$ are $c_1 = (3/2)^{2/3} \approx 1.31$ and $c_2 = 2$, respectively, so that the value expansion in Theorem 3.3 reduces to the corresponding results for proportional costs [2] and quadratic costs [50] in these cases. For $p \in (1, 2)$, it needs to be computed numerically. It turns out that $p \mapsto c_p$ is increasing; for the empirically most relevant case of $p \approx 3/2$, we have $c_p \approx 1.76$; cf. [32], Figure 3.

The final ingredient for the value expansion is the average of the other model parameters. In the Black–Scholes model of Guasoni and Weber [32], this term is constant. In the general model considered here, all these quantities are stochastic processes and, therefore, need to be averaged appropriately both with respect to time and states. Like for proportional and quadratic costs [2, 43, 50], the averaging with respect to states is performed with respect to the frictionless minimal entropy martingale measure $\hat{\mathbb{Q}}$. In view of [18], this means that the effect of the small friction is priced like a "marginal" path-dependent option. Like for other trading costs [7, 43, 50], this price is determined by (i) the trading cost, (ii) the volatility of the risky asset, (iii) the volatility of the frictionless target strategy and (iv) the agent's risk aversion. The powers through which these quantities enter interpolate between the cases of proportional and quadratic costs. The comparative statics are the same in each case: the transaction costs cause a big welfare effect if (i) trading costs are large, (ii) the risky asset is volatile necessitating close tracking of the optimal risk-return allocation, (iii) the frictionless target is volatile so that its tracking leads to substantial trading costs and (iv) risk aversion is high so that displacements from the optimal risk-return tradeoff have a big effect.

3.4. *Examples and applications.* Let us now discuss some examples and applications for our main result, Theorem 3.3. More specifically, we sketch how it can be used to study the effects of nonlinear trading costs in two of the settings where they are of crucial importance: the pricing and hedging of derivative securities, as well as active portfolio management.

Hedging of derivatives. Let us first illustrate how to use Theorem 3.3 to implement hedging strategies in the presence of small nonlinear price impact. For concreteness, we consider a Bachelier model with dynamics

$$dS_t = \sigma \, dW_t^{\mathbb{P}}.$$

Here, $W^{\mathbb{P}}$ is a standard Brownian motion, and σ is a positive constant. Let us study the optimization problem of an agent that has sold a European option with payoff function $H = h(S_T)$, where $s \mapsto h(s)$ is four times differentiable and with bounded derivatives and, additionally, h'' is bounded away from zero.[16] Then H is replicated by the delta hedge $\partial_s f(t, S_t)$, where the option price $f(t, S_t)$ at time t is

$$f(t, s) = \int_{-\infty}^{\infty} h(s + x\sigma\sqrt{T-t})\phi(x)\, dx.$$

[16] These regularity conditions parallel those generally required for proportional [1, 11], fixed [22] or quadratic costs [50]. With substantial additional effort, the case of a put option is worked out in a model with fixed costs in [22].

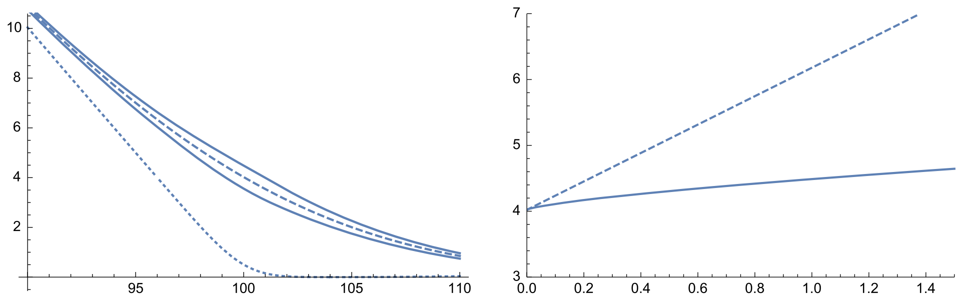

FIG. 1. *Left panel: payoff of the smooth concavified put (dotted), its frictionless Bachelier price (dashed) and its indifference prices for buying and selling one claim when hedging is subject to nonlinear price impact (solid) plotted against the initial price of the risky asset. Right panel: indifference price per claim plotted against the number of smooth, concavified puts traded, with nonlinear price impact (solid) and linear price impact (dashed).*

(Here, ϕ denotes the density function of the standard normal law.) Dominated convergence shows that f is four times differentiable with bounded derivatives. Moreover, the option's "gamma" $\Gamma_t = \partial_{ss} f(t, S_t)$ is also bounded away from zero.

Now, note that Jensen's inequality and the $\mathbb{P}$-martingale property of admissible strategies in the present context show that the replicating strategy $\partial_s f(t, S_t)$ is optimal for the utility maximization problem (2.4) augmented by the short position in H. This problem is equivalent to the optimization problem without the claim H under the measure $\mathbb{P}^H$ with density $d\mathbb{P}^H/d\mathbb{P} = e^{-\gamma H}/\mathbb{E}_{\mathbb{P}}[e^{-\gamma H}]$. Whence, we can apply Theorem 3.3 for constant positive Λ, for example: Assumption 1 then holds with $\hat{\mathbb{Q}} = \mathbb{P}$, Assumption 2 is evidently satisfied, and Assumption 3 also holds because the (constant) diffusion coefficient of the risky asset as well as the frictionless target strategy and its drift and diffusion coefficients are bounded, and because $c^{\hat{\varphi}}$ (and, therefore, m) is bounded away from 0. The asymptotically optimal trading rate is in turn given by (3.6); the corresponding performance loss is given by the formula from Theorem 3.3. For exponential utility, the certainty equivalent loss obtained by disregarding the frictionless Lagrange multiplier $\hat{y}$ corresponds to the adjustment of the utility-indifference price of Hodges and Neuberger [36]; compare [11, 42, 61].

As a concrete example, let us consider a "smoothed convexified put option" with strike K. Here, smoothing refers to replacing the actual put payoff $(K - S_T)^+$ with its Bachelier price $(K - S_T)\Phi(\frac{K-S_T}{\sigma\sqrt{\vartheta}}) + \sigma\sqrt{\vartheta}\phi(\frac{K-S_T}{\sigma\sqrt{\vartheta}})$ with a very short maturity ϑ, say one day. This ensures that the payoff has bounded smooth derivatives of all orders. Since the second derivative of this payoff and in turn the diffusion coefficient of the replicating strategy is not bounded away from zero, we slightly modify the payoff further by smoothly adding suitable parabolas for sufficiently large and small values of the terminal asset price. The resulting payoff function then satisfies all assumptions made above; it is depicted in the left panel of Figure 1. There, we also plot the corresponding Bachelier price and the illiquidity corrections derived from Theorem 3.3 by numerical integration for a long and a short position of one option, respectively. As (yearly) parameters, we use $\sigma = 0.2 \times 100$, which roughly corresponds to a Black–Scholes volatility of 20% at initial price 100 (compare [59]), $\gamma = 10$, $p = 3/2$ and the estimate $\lambda = 0.14 \times 1.57/250$ from [6].

To illustrate the nonlinear scaling induced by the nonlinear price impact, the right panel in Figure 1 plots the corresponding liquidity-adjusted price per claim for various numbers of an at-the-money, smooth concavified put. The resulting nonlinear prices are compared to their counterparts in a model with linear price impact $p = 2$, keeping all other parameters the same.[17] Clearly, the prices with elasticity $p = 3/2$ are higher only for very small trade sizes; for the larger trades necessary to hedge larger positions, the adjustments with linear costs quickly become bigger.

Active portfolio management. We now turn to a portfolio-choice model where randomly changing investment opportunities lead to active portfolio management. To wit, we consider the Kim–Omberg model [44] with mean-reverting returns:

$$dS_t = \mu_t \, dt + \sigma \, dW_t^{\mathbb{P}}, \tag{3.13}$$

where μ_t is an Ornstein–Uhlenbeck process:

$$d\mu_t = \kappa(\bar{\mu} - \mu_t)\,dt + \sigma_\mu \, dZ_t^{\mathbb{P}}.$$

Here, σ, $\bar{\mu}$, κ, σ_μ are positive constants and $W^{\mathbb{P}}$, $Z^{\mathbb{P}}$ are standard $\mathbb{P}$-Brownian motions with constant correlation $\rho \leq 0$.[18] In this setting, Assumption 1 is satisfied, and the frictionless optimal portfolio is [44]:

$$\hat{\varphi}_t = \frac{\mu_t}{\gamma\sigma^2} + \frac{\rho\sigma_\mu}{\gamma\sigma}\big(C(t)\mu_t + B(t)\big),$$

for nonpositive, smooth functions $C(t)$, $B(t)$ solving some Riccati equations.[19] Thus, $c_t^S = \sigma^2$ and $c_t^{\hat{\varphi}} = \frac{\sigma_\mu^2}{\gamma^2\sigma^4}(1 + \rho\sigma\sigma_\mu C(t))^2$ are deterministic, bounded and bounded away from zero here. Furthermore, B and C are continuous, so Assumption 2 holds.

Next note that $\hat{\varphi}_t$ is the sum of an Ornstein–Uhlenbeck process (with bounded, time-dependent mean-reversion level and speed) and a bounded function under the minimal entropy measure $\hat{\mathbb{Q}}$. Whence, its supremum has finite $\hat{\mathbb{Q}}$-moments of all orders. Therefore, the moment conditions in Assumption 3 are satisfied for constant Λ, for example. The exponential moment conditions also hold if the time horizon is sufficiently short.[20] The asymptotically optimal trading rate (3.6) in turn is a deterministic function of the deviation from the frictionless target, similarly as in the Black–Scholes model of [32].

In the uncorrelated case ($\rho = 0$), the relative trading rate is constant, in line with the constant relative trading rates of [26, 27] and the no-trade regions of constant width in [19, 49]. The corresponding certainty equivalent loss from Theorem 3.3 then also accumulates at a constant rate:

$$\lambda^{\frac{2}{2+p}} \left(\frac{\sigma_\mu^2}{8\gamma\sigma^2}\right)^{\frac{p}{p+2}} c_p.$$

The performance loss therefore is increasing in (i) the trading cost, (ii) the volatility of the signal μ_t and (iii) the inverse of the product of risk-aversion and asset volatility. The intuition for the last scaling is that the frictionless target is also inversely proportional to this term, and the resulting reduction of the frictionless target volatility overrides the increase of the tracking speed. In contrast, the mean-reversion level $\bar{\mu}$ and mean-reversion speed κ of the expected returns do not influence the leading-order term.

---

[17] In particular, we use the same value of λ as in [6]. Ideally, this scaling parameter of course should be estimated for each elasticity of price impact p from the same dataset, but such estimates do not seem to be available in the literature.

[18] Empirical studies such as [9] typically find substantially negative values. For the uncorrelated case ($\rho = 0$), the optimal portfolio is the same as for the local mean-variance criteria of [26, 27, 40, 49].

[19] Assumption 1 follows from [48], Example 3 in Section 6.2; admissibility can be established using [54], Lemma 2.12, and [48], Example 3 in Section 6.2.

[20] This restriction could be avoided by either directly working with quadratic rather than exponential preferences as in [27], or by truncating large values of the state variable as in [50], Section 8.1.

4. Primal considerations. To prove Theorem 3.3, we first establish that the expected utility corresponding to the candidate strategy φ^λ from (3.11) can be bounded from below by the asymptotic expansion asserted in Theorem 3.3.

As a preliminary observation, we note that φ^λ is admissible and in $L^2_{\mathbb{Q}}(S)$. Indeed, the elementary inequality $|x| \le 1 + |x|^{1+\varepsilon}$ for $\varepsilon > 0$, Lemma B.1 and Jensen's inequality give

$$\begin{aligned}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T (\varphi_t^\lambda)^2\, d\langle S\rangle_t\right] &\le 1 + \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_0^T (\varphi_t^\lambda)^2 c_t^S\, dt\right)^{1+\varepsilon}\right]\\ &\le 1 + T^\varepsilon \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T ((\hat{\varphi}_t^*)^2 c_t^S)^{1+\varepsilon}\, dt\right] < \infty.\end{aligned}$$

4.1. *Approximation by cost-displacement tradeoff.* We start our analysis by a Taylor expansion of the utility of the frictional candidate wealth process around its frictionless counterparts. Together with careful remainder estimates, this shows that the expected utility loss when applying the strategy φ^λ from (3.11) can be asymptotically bounded from below by a tradeoff between squared displacement from the frictionless target and accumulated trading costs.[21]

PROPOSITION 4.1. *Suppose Assumptions 1, 2 and 3 are satisfied. Then*

$$\begin{aligned}&J^\lambda(\varphi^\lambda) - J^0(\hat{\varphi})\\ &\quad \ge -\hat{y}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}} \left(\frac{\gamma}{2}(\Delta\varphi_t^\lambda)^2 c_t^S + \lambda\Lambda_t|\dot{\varphi}_t^\lambda|^p\right) dt\right] + o(\lambda^{\frac{2}{p+2}}).\end{aligned} \tag{4.1}$$

PROOF. To ease notation, set $V_T = x + \int_0^T \hat{\varphi}_t\, dS_t$, $V_T^\lambda = x + \int_0^T \varphi_t^\lambda\, dS_t - \lambda\int_0^T \Lambda_t \times |\dot{\varphi}_t^\lambda|^p\, dt$, as well as $\bar{V}_T^\lambda = x + \int_0^T \varphi_t^\lambda\, dS_t - \lambda\int_0^{\tau^{\Delta\varphi}} \Lambda_t|\dot{\varphi}_t^\lambda|^p\, dt$.

Step 1. We first establish some preliminary estimates. To this end, set

$$\begin{aligned}A_1 &:= \int_0^{\tau^{\Delta\varphi}} \Delta\varphi_t^\lambda\, dS_t, &\qquad A_2 &:= \int_{\tau^{\Delta\varphi}}^T \Delta\varphi_t^\lambda\, dS_t,\\ A_3 &:= \int_0^{\tau^{\Delta\varphi}} \lambda\Lambda_t|\dot{\varphi}_t^\lambda|^p\, dt, &\qquad A_4 &:= \int_{\tau^{\Delta\varphi}}^T \lambda\Lambda_t|\dot{\varphi}_t^\lambda|^p\, dt.\end{aligned}$$

Note that under the martingale measure $\hat{\mathbb{Q}}$, the SDE satisfied by the asset price S is

$$dS_t = \sqrt{c_t^S}\, dW_t^{S,\hat{\mathbb{Q}}},$$

where $W^{S,\hat{\mathbb{Q}}}$ is a $\hat{\mathbb{Q}}$ Brownian motion. The Itô isometry and Proposition 4.4 below give

$$\mathbb{E}_{\hat{\mathbb{Q}}}[(A_1)^2] = \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}} (\Delta\varphi_t^\lambda)^2 c_t^S\, dt\right] = O(\lambda^{\frac{2}{p+2}}), \tag{4.2}$$

$$\mathbb{E}_{\hat{\mathbb{Q}}}[A_3] = O(\lambda^{\frac{2}{p+2}}). \tag{4.3}$$

Moreover, Itô's isometry, the Burkholder–Davis–Gundy inequality (with constant C_{BDG}), and Lemma 4.3 show that

$$\mathbb{E}_{\hat{\mathbb{Q}}}[(A_2)^2] = \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_{\tau^{\Delta\varphi}}^T (\Delta\varphi_t^\lambda)^2 c_t^S\, dt\right] = o(\lambda^{\frac{2}{p+2}}), \tag{4.4}$$

[21] Similar goal functionals are directly used in a number of studies; cf., for example, [5, 49]. Related pathwise criteria are studied in [12, 28, 56].

$$\mathbb{E}_{\hat{\mathbb{Q}}}\big[(A_2)^6\big] \leq C_{\mathrm{BDG}}\mathbb{E}_{\hat{\mathbb{Q}}}\bigg[\bigg(\int_{\tau^{\Delta\varphi}}^T (\Delta\varphi_t^\lambda)^2 c_t^S\,dt\bigg)^3\bigg] = o\big(\lambda^{\frac{4}{p+2}}\big). \tag{4.5}$$

Furthermore, by the Burkholder–Davis–Gundy inequality, the definition of $\tau^{\Delta\varphi}$ in (3.10), the choices of κ_1, κ_3, κ_4, and the integrability conditions from Assumption 3, we obtain

$$\mathbb{E}_{\hat{\mathbb{Q}}}\big[(A_1)^6\big] \leq C_{\mathrm{BDG}} T^3 \lambda^{6\kappa_1}\mathbb{E}_{\hat{\mathbb{Q}}}\big[(c_T^{S*})^3\big] = o\big(\lambda^{\frac{4}{p+2}}\big) \tag{4.6}$$

$$A_3 \leq T\lambda^{\kappa_3} = o\big(\lambda^{\frac{1}{3}\frac{2}{p+2}}\big), \tag{4.7}$$

$$\begin{aligned} A_4 &\leq \int_{\tau^{\Delta\varphi}}^{\tau^{\Delta\varphi}+\lambda^\eta} \lambda\Lambda_t\big|\lambda^{-\eta}\varphi_{\tau^{\Delta\varphi}}^\lambda\big|^p\,dt \leq \lambda^{1-(p-1)\eta-p\kappa_4}\Lambda_T^* \\ &= o\big(\lambda^{\frac{2}{p+2}}\big)\Lambda_T^*. \end{aligned} \tag{4.8}$$

Combining (4.3) and (4.7), we now deduce

$$\mathbb{E}_{\hat{\mathbb{Q}}}\big[(A_3)^2\big] = o\big(\lambda^{\frac{1}{3}\frac{2}{p+2}}\big)\mathbb{E}_{\hat{\mathbb{Q}}}[A_3] = o\big(\lambda^{\frac{4}{3}\frac{2}{p+2}}\big). \tag{4.9}$$

Step 2. Next, we show that

$$\mathbb{E}_{\mathbb{P}}\big[U\big(V_T^\lambda\big)\big] \geq \mathbb{E}_{\mathbb{P}}\big[U\big(\bar{V}_T^\lambda\big)\big] + o\big(\lambda^{\frac{2}{p+2}}\big). \tag{4.10}$$

By concavity of U, the fact that U is an exponential function, the frictionless first-order condition $U'(V_T) = \hat{y}\,d\hat{\mathbb{Q}}/d\mathbb{P}$, the identities $V_T - V_T^\lambda = A_1 + A_2 + A_3 + A_4$ and $V_T^\lambda - \bar{V}_T^\lambda = -A_4$, as well as the estimates (4.7), (4.8), we obtain for $\lambda \leq 1$,

$$\begin{aligned} &\mathbb{E}_{\mathbb{P}}\big[U\big(V_T^\lambda\big)\big] - \mathbb{E}_{\mathbb{P}}\big[U\big(\bar{V}_T^\lambda\big)\big] \\ &\quad\geq \mathbb{E}_{\mathbb{P}}\big[U'\big(V_T^\lambda\big)\big(V_T^\lambda - \bar{V}_T^\lambda\big)\big] = \hat{y}\mathbb{E}_{\hat{\mathbb{Q}}}\big[\exp\big(-\gamma\big(V_T^\lambda - V_T\big)\big)\big(V_T^\lambda - \bar{V}_T^\lambda\big)\big] \\ &\quad\geq -\hat{y}\exp\big(\gamma T\lambda^{\kappa_3}\big)o\big(\lambda^{\frac{2}{p+2}}\big) \\ &\qquad\times \mathbb{E}_{\hat{\mathbb{Q}}}\bigg[\exp\bigg(\bigg|\int_0^T \gamma\,\Delta\varphi_t^\lambda\,dS_t\bigg| + \gamma\lambda^{\frac{2}{2+p}}\Lambda_T^*\bigg)\Lambda_T^*\bigg]. \end{aligned} \tag{4.11}$$

To establish (4.10), it suffices to show that the expectation on the right-hand side of (4.11) is of order $O(1)$. To this end, let λ be sufficiently small and choose ε as in Assumption 3. Then the elementary inequality $x \leq \frac{2(1+\varepsilon)}{\varepsilon^2}\exp(\frac{\varepsilon^2}{2(1+\varepsilon)}x)$ for $x \geq 0$, Hölder's inequality with exponents $1+\varepsilon$ and $1+\frac{1}{\varepsilon}$ and Lemma 4.2 give

$$\begin{aligned} &\mathbb{E}_{\hat{\mathbb{Q}}}\bigg[\exp\bigg(\bigg|\int_0^T \gamma\,\Delta\varphi_t^\lambda\,dS_t\bigg| + \gamma\lambda^{\frac{2}{2+p}}\Lambda_T^*\bigg)\Lambda_T^*\bigg] \\ &\quad\leq \mathbb{E}_{\hat{\mathbb{Q}}}\bigg[\exp\bigg(\bigg|\int_0^T \gamma\,\Delta\varphi_t^\lambda\,dS_t\bigg|\bigg)\exp\bigg(\frac{\varepsilon^2}{2(1+\varepsilon)}\Lambda_T^*\bigg)\Lambda_T^*\bigg] \\ &\quad\leq \frac{2(1+\varepsilon)}{\varepsilon^2}\mathbb{E}_{\hat{\mathbb{Q}}}\bigg[\exp\bigg(\bigg|\int_0^T \gamma\,\Delta\varphi_t^\lambda\,dS_t\bigg|\bigg)\exp\bigg(\frac{\varepsilon^2}{(1+\varepsilon)}\Lambda_T^*\bigg)\bigg] \\ &\quad\leq \frac{2(1+\varepsilon)}{\varepsilon^2}\mathbb{E}_{\hat{\mathbb{Q}}}\bigg[\exp\bigg((1+\varepsilon)\gamma\bigg|\int_0^T \Delta\varphi_t^\lambda\,dS_t\bigg|\bigg)\bigg]^{\frac{1}{1+\varepsilon}}\mathbb{E}_{\hat{\mathbb{Q}}}\big[\exp\big(\varepsilon\Lambda_T^*\big)\big]^{\frac{\varepsilon}{1+\varepsilon}} \\ &\quad\leq \frac{4(1+\varepsilon)}{\varepsilon^2}\mathbb{E}_{\hat{\mathbb{Q}}}\bigg[\exp\bigg(8(1+\varepsilon)^2\gamma^2\int_0^T (\hat{\varphi}_t^*)^2 c_t^S\,dt\bigg)\bigg]^{\frac{1}{(1+\varepsilon)}}\mathbb{E}_{\hat{\mathbb{Q}}}\big[\exp\big(\varepsilon\Lambda_T^*\big)\big]^{\frac{\varepsilon}{1+\varepsilon}} \\ &\quad< \infty. \end{aligned}$$

Thus, (4.10) indeed holds as asserted.

Step 3. Finally, we show that

$$\begin{aligned}&\mathbb{E}_{\mathbb{P}}\big[U\big(\bar{V}_T^{\lambda}\big)\big]-\mathbb{E}_{\mathbb{P}}\big[U(V_T)\big]\\&\quad\geq -\hat{y}\mathbb{E}_{\hat{\mathbb{Q}}}\bigg[\int_0^{\tau^{\Delta\varphi}}\bigg(\frac{\gamma}{2}\big(\Delta\varphi_t^{\lambda}\big)^2c_t^S+\lambda\Lambda_t\big|\dot{\varphi}_t^{\lambda}\big|^p\bigg)\,dt\bigg]+o\big(\lambda^{\frac{2}{p+2}}\big).\end{aligned}\tag{4.12}$$

Together with (4.10), this yields the asymptotic lower bound (4.1). For fixed ω, a second-order Taylor expansion with Lagrange remainder term of $U(\bar{V}_T^{\lambda})$ around the frictionless optimizer V_T yields

$$\begin{aligned}U\big(\bar{V}_T^{\lambda}\big)&=U(V_T)+U'(V_T)\big(\bar{V}_T^{\lambda}-V_T\big)+\frac{1}{2}U''(V_T)\big(\bar{V}_T^{\lambda}-V_T\big)^2\\&\quad+\frac{1}{6}U'''\big(\xi(\omega)\big)\big(\bar{V}_T^{\lambda}-V_T\big)^3,\end{aligned}$$

where $\xi(\omega)$ takes values in the interval with endpoints $V_T(\omega)$ and $\bar{V}_T^{\lambda}(\omega)$. Using that $U''/U'\equiv-\gamma$, $U'''/U'\equiv\gamma^2$, and U' is positive, we obtain

$$\begin{aligned}U\big(\bar{V}_T^{\lambda}\big)-U(V_T)&\geq -U'(V_T)\bigg(\big(V_T-\bar{V}_T^{\lambda}\big)+\frac{1}{2}\gamma\big(V_T-\bar{V}_T^{\lambda}\big)^2\\&\qquad+\frac{1}{6}\gamma^2\exp\big(\gamma\big|V_T-\bar{V}_T^{\lambda}\big|\big)\big|V_T-\bar{V}_T^{\lambda}\big|^3\bigg).\end{aligned}\tag{4.13}$$

By the first-order condition $U'(V_T)=\hat{y}\,d\hat{\mathbb{Q}}/d\mathbb{P}$ and Bayes' theorem, it remains to show that

$$\begin{aligned}&\mathbb{E}_{\hat{\mathbb{Q}}}\bigg[\big(V_T-\bar{V}_T^{\lambda}\big)+\frac{1}{2}\gamma\big(V_T-\bar{V}_T^{\lambda}\big)^2+\frac{1}{6}\gamma^2\exp\big(\gamma\big|V_T-\bar{V}_T^{\lambda}\big|\big)\big|V_T-\bar{V}_T^{\lambda}\big|^3\bigg]\\&\quad=\mathbb{E}_{\hat{\mathbb{Q}}}\bigg[\int_0^{\tau^{\Delta\varphi}}\bigg(\frac{\gamma}{2}\big(\Delta\varphi_t^{\lambda}\big)^2c_t^S+\lambda\Lambda_t\big|\dot{\varphi}_t^{\lambda}\big|^p\bigg)\,dt\bigg]+o\big(\lambda^{\frac{2}{p+2}}\big).\end{aligned}\tag{4.14}$$

We split up the expectation on the left-hand side of (4.14) into three parts and use that $V_T-\bar{V}_T^{\lambda}=A_1+A_2+A_3$. First, using that that $\int_0^{\cdot}\varphi^{\lambda}\,dS$ and $\int_0^{\cdot}\hat{\varphi}\,dS$ are $\hat{\mathbb{Q}}$-martingales (cf. Theorem 3.3), we obtain

$$\begin{aligned}\mathbb{E}_{\hat{\mathbb{Q}}}\big[V_T-\bar{V}_T^{\lambda}\big]&=\mathbb{E}_{\hat{\mathbb{Q}}}[A_1+A_2+A_3]=\mathbb{E}_{\hat{\mathbb{Q}}}[A_3]\\&=\mathbb{E}_{\hat{\mathbb{Q}}}\bigg[\int_0^{\tau^{\Delta\varphi}}\lambda\Lambda_t\big|\dot{\varphi}_t^{\lambda}\big|^p\,dt\bigg].\end{aligned}\tag{4.15}$$

Next, the Cauchy–Schwarz inequality and the estimates (4.2), (4.4), (4.9) give

$$\begin{aligned}&\mathbb{E}_{\hat{\mathbb{Q}}}\big[\big(V_T-\bar{V}_T^{\lambda}\big)^2\big]\\&\quad=\mathbb{E}_{\hat{\mathbb{Q}}}\big[A_1^2+2A_1A_2+2A_1A_3+2A_2A_3+A_2^2+A_3^2\big]\\&\quad\leq\mathbb{E}_{\hat{\mathbb{Q}}}\big[A_1^2\big]+2\mathbb{E}_{\hat{\mathbb{Q}}}\big[A_1^2\big]^{\frac{1}{2}}\mathbb{E}_{\hat{\mathbb{Q}}}\big[A_2^2\big]^{\frac{1}{2}}+2\mathbb{E}_{\hat{\mathbb{Q}}}\big[A_1^2\big]^{\frac{1}{2}}\mathbb{E}_{\hat{\mathbb{Q}}}\big[A_3^2\big]^{\frac{1}{2}}\\&\qquad+2\mathbb{E}_{\hat{\mathbb{Q}}}\big[A_2^2\big]^{\frac{1}{2}}\mathbb{E}_{\hat{\mathbb{Q}}}\big[A_3^2\big]^{\frac{1}{2}}+\mathbb{E}_{\hat{\mathbb{Q}}}\big[A_2^2\big]+\mathbb{E}_{\hat{\mathbb{Q}}}\big[A_3^2\big]\\&\quad\leq\mathbb{E}_{\hat{\mathbb{Q}}}\big[A_1^2\big]+2O\big(\lambda^{\frac{1}{p+2}}\big)o\big(\lambda^{\frac{1}{p+2}}\big)+2O\big(\lambda^{\frac{1}{p+2}}\big)o\big(\lambda^{\frac{2}{3}\frac{2}{p+2}}\big)\\&\qquad+2o\big(\lambda^{\frac{1}{p+2}}\big)o\big(\lambda^{\frac{2}{3}\frac{2}{p+2}}\big)+o\big(\lambda^{\frac{2}{p+2}}\big)+o\big(\lambda^{\frac{4}{3}\frac{2}{p+2}}\big)\\&\quad=\mathbb{E}_{\hat{\mathbb{Q}}}\bigg[\int_0^{\tau^{\Delta\varphi}}\big(\Delta\varphi_t^{\lambda}\big)^2c_t^S\,dt\bigg]+o\big(\lambda^{\frac{2}{p+2}}\big).\end{aligned}\tag{4.16}$$

Finally, using the elementary inequality $(a+b+c)^3 \leq 9(a^3+b^3+c^3)$ for $a,b,c \geq 0$, the Cauchy–Schwarz inequality, the estimates (4.6), (4.5), (4.7) and Lemma 4.2, we obtain

$$\begin{aligned}
&\mathbb{E}_{\hat{\mathbb{Q}}}[\exp(\gamma|V_T - \bar{V}_T^\lambda|)|V_T - \bar{V}_T^\lambda|^3] \\
&\quad \leq 9\exp(\gamma T\lambda^{\kappa_3})\mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(\left|\int_0^T \gamma\Delta\varphi_t^\lambda\, dS_t\right|\right)((A_1)^3+(A_2)^3+(A_3)^3)\right] \\
&\quad \leq 9\exp(\gamma T\lambda^{\kappa_3})\mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(\left|\int_0^T 2\gamma\Delta\varphi_t^\lambda\, dS_t\right|\right)\right]^{\frac{1}{2}} \\
&\qquad \times (\mathbb{E}_{\hat{\mathbb{Q}}}[(A_1)^6]^{\frac{1}{2}} + \mathbb{E}_{\hat{\mathbb{Q}}}[(A_2)^6]^{\frac{1}{2}}) \\
&\qquad + 9\exp(\gamma T\lambda^{\kappa_3})o(\lambda^{\frac{2}{p+2}})\mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(\left|\int_0^T \gamma\Delta\varphi_t^\lambda\, dS_t\right|\right)\right] \\
&\quad = o(\lambda^{\frac{2}{p+2}}).
\end{aligned} \tag{4.17}$$

Combining (4.15)–(4.17) yields (4.14), thereby completing the proof. □

The following two auxiliary estimates are used in the proof of Proposition 4.1.

LEMMA 4.2. *Suppose Assumptions* 1, 2 *and* 3 *are satisfied. Then, for* $k \in [1,2]$,

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(k\gamma\left|\int_0^T \Delta\varphi_t^\lambda\, dS_t\right|\right)\right] \leq 2\mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(8k^2\gamma^2\int_0^T (\hat{\varphi}_t^*)^2 c_t^S\, dt\right)\right] < \infty.$$

PROOF. Fix $k \in [1,2]$. In view of the elementary inequality $\exp(|x|) \leq \exp(x) + \exp(-x)$ for $x \in \mathbb{R}$, it suffices to show that

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(\int_0^T \pm k\gamma\Delta\varphi_t^\lambda\, dS_t\right)\right] \leq \mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(8k^2\gamma^2\int_0^T (\hat{\varphi}_t^*)^2 c_t^S\, dt\right)\right] < \infty.$$

This follows from [53], Theorem III.43, and Lemma B.1. □

LEMMA 4.3. *Suppose Assumptions* 1, 2 *and* 3 *are satisfied. Then for* $k \in [1,3]$,

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_{\tau^{\Delta\varphi}}^T (\Delta\varphi_t^\lambda)^2 c_t^S\, dt\right)^k\right] = o(\lambda^{\frac{2\min(k,2)}{p+2}}).$$

PROOF. Fix $k \in [1,3]$. In view of the elementary inequality $(a+b)^k \leq 4(a^k+b^k)$ for $a,b \geq 0$, it suffices to show that

$$\begin{aligned}
&\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_{\tau^{\Delta\varphi}}^{T^\lambda} (\Delta\varphi_t^\lambda)^2 c_t^S\, dt\right)^k\right] = o(\lambda^{\frac{4}{p+2}}) \quad \text{and} \\
&\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_{T^\lambda}^T (\Delta\varphi_t^\lambda)^2 c_t^S\, dt\right)^k\right] = o(\lambda^{\frac{2k}{p+2}}).
\end{aligned} \tag{4.18}$$

To establish the first part of (4.18), we use Lemma B.1, Hölder's inequality, and Proposition C.5 to obtain

$$\begin{aligned}
&\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_{\tau^{\Delta\varphi}}^{T^\lambda} (\varphi_t^\lambda - \hat{\varphi}_t)^2 c_t^S\, dt\right)^k\right] \\
&\quad \leq 2^{2k}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_{\tau^{\Delta\varphi}}^{T^\lambda} (\hat{\varphi}_t^*)^2 c_t^S\, dt\right)^k \mathbb{1}_{\{\tau^{\Delta\varphi} < T^\lambda\}}\right] \\
&\quad \leq 2^{2k}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_0^T (\hat{\varphi}_t^*)^2 c_t^S\, dt\right)^{k(1+\frac{1}{\varepsilon})}\right]^{\frac{1}{1+\frac{1}{\varepsilon}}} \hat{\mathbb{Q}}[\tau^{\Delta\varphi} < T^\lambda]^{\frac{1}{1+\varepsilon}} = o(\lambda^{\frac{4}{p+2}}).
\end{aligned}$$

To establish the second part of (4.18), we use Lemma B.1, $T - T^\lambda = \lambda^\eta$, and $\eta > \frac{2}{p+2}$ to obtain

$$\mathbb{E}_{\hat{\mathbb{Q}}}\Big[\Big(\int_{T^\lambda}^T (\varphi_t^\lambda - \hat{\varphi}_t)^2 c_t^S\, dt\Big)^k\Big] \leq 2^{2k}\lambda^{k\eta}\mathbb{E}_{\hat{\mathbb{Q}}}\big[\big((\hat{\varphi}_T^*)^2 c_T^{S*}\big)^k\big] = O(\lambda^{k\eta}) = o(\lambda^{\frac{2k}{p+2}}).$$

This completes the proof. □

4.2. *Computation of the cost-displacement tradeoff.* In Section 4.1, we have seen that the exponential utility generated by the candidate strategy from Section 3.1 is asymptotically bounded from below by a tradeoff between squared displacement from the frictionless target and accumulated trading costs. By applying ergodic results developed in the companion paper of the present study [15], this cost-displacement tradeoff can be computed explicitly in terms of the model parameters.

PROPOSITION 4.4. *Suppose Assumptions 1, 2 and 3 are satisfied. Then*

$$\begin{aligned}
&\mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_0^{\tau^{\Delta\varphi}} \Big(\frac{\gamma}{2}(\Delta\varphi_t^\lambda)^2 c_t^S + \lambda_t|\dot{\varphi}_t^\lambda|^p\Big)\, dt\Big] \\
&\quad = \lambda^{\frac{2}{p+2}} c_p \mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_0^T \Lambda_t^{\frac{2}{p+2}}\big(2^{-3}\gamma c_t^S (c_t^{\hat{\varphi}})^2\big)^{\frac{p}{p+2}}\, dt\Big] + o(\lambda^{\frac{2}{p+2}}) \\
&\quad = O(\lambda^{\frac{2}{p+2}}),
\end{aligned} \tag{4.19}$$

where c_p is the constant defined in Lemma 3.1.

PROOF. To ease notation, set

$$C(\varphi^\lambda) := \mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_0^{\tau^{\Delta\varphi}} \Big(\frac{\gamma}{2}(\Delta\varphi_t^\lambda)^2 c_t^S + \lambda_t|\dot{\varphi}_t^\lambda|^p\Big)\, dt\Big]. \tag{4.20}$$

The stopping time $\tau^{\Delta\varphi}$ converges to T in probability as λ goes to zero by Proposition C.5. By the dominated convergence theorem and Assumption 3, it therefore suffices to show that

$$C(\varphi^\lambda) = \lambda^{\frac{2}{p+2}} c_p \mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_0^{\tau^{\Delta\varphi}} \Lambda_t^{\frac{2}{p+2}}\big(2^{-3}\gamma c_t^S (c_t^{\hat{\varphi}})^2\big)^{\frac{p}{p+2}}\, dt\Big] + o(\lambda^{\frac{2}{p+2}}). \tag{4.21}$$

By the definition (3.11) of $\Delta\varphi^\lambda$ and since $\Delta\varphi^\lambda$ coincides with $\overline{\Delta\varphi}^\lambda$ on $[\![0, \tau^\Delta]\!]$, we have

$$\begin{aligned}
C(\varphi^\lambda) = \mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_0^{\tau^{\Delta\varphi}} &\Big(\frac{\gamma}{2}(\overline{\Delta\varphi}_t^\lambda)^2 c_t^S \\
&+ 2^{-p} p^{-\frac{p}{p-1}} \lambda^{\frac{2}{p+2}} \Lambda_t (c_t^{\hat{\varphi}})^p m_t^p \big|\tilde{g}_p\big(\lambda^{-\frac{1}{p+2}} m_t \overline{\Delta\varphi}_t^\lambda\big)\big|^p\Big)\, dt\Big].
\end{aligned} \tag{4.22}$$

We now apply [15], Theorem 1.3, to the right-hand side of (4.22). To this end, set

$$v = \frac{\int_{\mathbb{R}} x^2 \exp(-p^{-\frac{1}{p-1}}\tilde{G}_p(x))\, dx}{\int_{\mathbb{R}} \exp(-p^{-\frac{1}{p-1}}\tilde{G}_p(x))\, dx}, \tag{4.23}$$

$$w = \frac{\int_{\mathbb{R}} |\tilde{g}_p(x)|^p \exp(-p^{-\frac{1}{p-1}}\tilde{G}_p(x))\, dx}{\int_{\mathbb{R}} \exp(-p^{-\frac{1}{p-1}}\tilde{G}_p(x))\, dx}, \tag{4.24}$$

where $\widetilde{G}_p(x) = \int_0^x \widetilde{g}_p(x)$ denotes the antiderivative of $\widetilde{g}_p$. Then apply (the $\mathcal{S}^1$-version of) [15], Theorem 1.3, with $X^\varepsilon = \overline{\Delta\varphi}^\lambda$, where $\varepsilon = \lambda^{\frac{1}{p+2}}$, $b = \mu^{\hat{\varphi},\hat{\mathbb{Q}}}$, $M = m$, $L = \frac{1}{2}p^{-\frac{1}{p-1}}m = \frac{1}{2}p^{-\frac{1}{p-1}}M$, and $c = c^{\hat{\varphi}}$, first to the case $f(x) = x^2$, $H = \frac{\gamma}{2}c^S$ and $K = 1$, and then to the case $f(x) = |\widetilde{g}_p(x)|^p$, $H = 2^{-p}p^{-\frac{p}{p-1}}\Lambda(c^{\hat{\varphi}})^p m^p$ and $K = m$. (Note that Assumption 3 ensures that [15], Assumptions 1, 2, are satisfied.[22]) This yields

$$C(\varphi^\lambda) = \lambda^{\frac{2}{p+2}}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}} v\frac{\gamma c_t^S}{2m_t^2} + p^{-\frac{p}{p-1}}w2^{-p}\Lambda_t(c_t^{\hat{\varphi}})^p m_t^p\,dt\right] + o(\lambda^{\frac{2}{p+2}}). \tag{4.25}$$

We proceed to simplify the expression inside the expectation of the right-hand side of (4.25). First, using the definition of m in (3.7), we obtain

$$\frac{\gamma c_t^S}{2m_t^2} = \Lambda_t^{\frac{2}{p+2}}\big(2^{-3}\gamma(c_t^{\hat{\varphi}})^2 c_t^S\big)^{\frac{p}{p+2}}, \tag{4.26}$$

$$2^{-p}\Lambda_t(c_t^{\hat{\varphi}})^p m_t^p = \Lambda_t^{\frac{2}{p+2}}\big(2^{-3}\gamma(c_t^{\hat{\varphi}})^2 c_t^S\big)^{\frac{p}{p+2}}. \tag{4.27}$$

Next, we show that

$$v + p^{-\frac{p}{p-1}}w = c_p. \tag{4.28}$$

To this end, notice that by Lemma A.1, the function $\widetilde{g}_p$ and its antiderivative $\widetilde{G}_p$ are bounded in norm (from above and below) by positive monomials of nonzero degree in a neighborhood of infinity. Hence, an integration by parts yields

$$\begin{aligned}
&\int_{\mathbb{R}} |\widetilde{g}_p(x)|^p \exp\big(-p^{-\frac{1}{p-1}}\widetilde{G}_p(x)\big)\,dx \\
&\quad = -p^{\frac{1}{p-1}}\int_{\mathbb{R}} |\widetilde{g}_p(x)|^{p-1}\operatorname{sgn}(x)\big(\exp\big(-p^{-\frac{1}{p-1}}\widetilde{G}_p(x)\big)\big)'\,dx \\
&\quad = -p^{\frac{1}{p-1}}\int_{\mathbb{R}} g_p(x)\big(\exp\big(-p^{-\frac{1}{p-1}}\widetilde{G}_p(x)\big)\big)'\,dx \\
&\quad = p^{\frac{1}{p-1}}\int_{\mathbb{R}} g_p'(x)\exp\big(-p^{-\frac{1}{p-1}}\widetilde{G}_p(x)\big)\,dx.
\end{aligned} \tag{4.29}$$

Plugging the ODE (3.1) for g_p into the right-hand side of (4.29), and then dividing both sides of (4.29) by $\int_{\mathbb{R}} \exp(-p^{-\frac{1}{p-1}}\widetilde{G}_p(x))\,dx$, we obtain

$$w = p^{\frac{1}{p-1}}\big((p-1)p^{-\frac{p}{p-1}}w - v + c_p\big).$$

Solving this equation for c_p yields the claimed relation (4.28).

The asserted formula (4.21) for the cost-displacement tradeoff now follows by plugging (4.26), (4.27), and (4.28) into (4.25). The last estimate follows from Hölder inequality and Assumption 3. □

5. Dual considerations.

5.1. *Asymptotic duality bound.* To complete the proof of Theorem 3.3, we now complement the primal lower bound from Proposition 4.1 with the following dual upper bound.

[22] More precisely, we have $p_{[15]} = 1$, $q_{[15]} = \frac{2}{p}$, $q'_{[15]} = 2$ for the constants of [15]. Assumption 1 of [15] is given by the integrability of the first item of $\mathcal{X}^\varepsilon$. The first two items of Assumption 2 in [15] are given by the second and third item of $\mathcal{X}^\varepsilon$. The third and fourth items of Assumption 2 in [15] are given by the fourth item of $\mathcal{X}^\varepsilon$ in the first case and by the fifth, sixth and seventh items of Assumption 3, Hölder's inequality, and the definition of $\kappa_1, \kappa_2, \kappa_3$ in the second case. The fifth item of Assumption 2 in [15] is given by the eighth item of $\mathcal{X}^\varepsilon$. The last two items of Assumption 2 in [15] are trivially satisfied as $\frac{L}{M}$ is a constant.

PROPOSITION 5.1. *Suppose Assumptions 1, 2 and 3 are satisfied. Then the following upper duality bound holds for any admissible strategy φ*:

$$J^\lambda(\varphi) - J^0(\hat{\varphi}) \le -\hat{y}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}} \left(\frac{\gamma}{2}(\Delta\varphi_t^\lambda)^2 c_t^S + \lambda_t|\dot{\varphi}_t^\lambda|^p\right) dt\right] + o(\lambda^{\frac{2}{p+2}}). \tag{5.1}$$

Suppose this result is established. Theorem 3.3 in turn follows from Proposition 5.1 combined with the primal lower bound established in Propositions 4.1 and 4.4 for the candidate strategy from Section 3.1.

5.2. *Proof of Proposition* 5.1. To prove Proposition 5.1, we use the duality theory for superlinear frictions developed very recently by Guasoni and Rasonyi [30]. They argue that in this context, the dual measures do not turn the frictionless price into a martingale, but rather the actual execution price with transaction costs.[23]

This characterization is apparently difficult to apply, since it relies on the primal and dual optimizers, both of which are unknown. However, it is extremely useful for the present asymptotic verification, because we already have a candidate *asymptotic* optimal strategy φ^λ at hand. We use the execution price corresponding to the latter as a substitute for the exact optimizer. However, this "naive" asymptotic execution price needs to be modified in two directions. First, as in the case of proportional transaction cost (cf. [1, 2, 34]) we need to stop the "naive" candidate in order to ensure enough integrability for the estimates in the remainder terms. Second, another subtle modification is necessary in order to control the displacement of the execution price from its frictionless counterpart; cf. Lemmas 5.4–5.8. This makes the analysis more delicate than in the case of proportional transaction costs. Indeed, unlike in [34], the change of measure from the frictionless martingale measure $\hat{\mathbb{Q}}$ to the frictional dual martingale measure $\hat{\mathbb{Q}}^\lambda$ (cf. (5.24) below) is no longer bounded so that more careful moment estimates are needed; cf. Lemmas 5.3 and 5.10.

Finally, with the modified asymptotic execution price and the corresponding frictionless martingale measure $\hat{\mathbb{Q}}^\lambda$ at hand, we use convex duality (both for the utility function and the trading cost functional), Taylor expansions and careful remainder estimates to derive the desired duality bound in Proposition 5.1.

"Naive" execution price. Recall that under the frictionless dual martingale measure $\hat{\mathbb{Q}}$, the risky asset has dynamics $dS_t = \sqrt{c_t^S}\, dW_t^{S,\hat{\mathbb{Q}}}$, where $W^{S,\hat{\mathbb{Q}}}$ is a $\hat{\mathbb{Q}}$-Brownian motion. Inspired by the first-order condition of [30], define the "naive" execution price for the candidate trading rate $\dot{\varphi}^\lambda$ from Section 3.1:

$$\overline{S}_t^\lambda := S_t + \overline{\Delta S}_t^\lambda := S_t + \lambda_t p \operatorname{sgn}(\dot{\varphi}_t^\lambda)|\dot{\varphi}_t^\lambda|^{p-1}. \tag{5.2}$$

Note that on $[\![0, \tau^{\Delta\varphi}]\!]$, we have

$$\overline{\Delta S}_t^\lambda = \lambda^{\frac{3}{p+2}} 2^{-(p-1)}\Lambda_t(c_t^{\hat{\varphi}})^{p-1} m_t^{p-1} g_p(\lambda^{-\frac{1}{p+2}} m_t \overline{\Delta\varphi}_t^\lambda) = \lambda^{\frac{3}{p+2}} A_t g_p(B_t^\lambda), \tag{5.3}$$

where

$$A_t = 2^{-(p-1)}\Lambda_t(c_t^{\hat{\varphi}})^{p-1} m_t^{p-1} \quad \text{and} \quad B_t^\lambda := \lambda^{-\frac{1}{p+2}} m_t \overline{\Delta\varphi}_t^\lambda. \tag{5.4}$$

[23] For models with proportional transaction costs, this leads to a "shadow price" in the spirit of [17, 38] that coincides with the bid- or ask-price, respectively, whenever the optimal strategy sells or purchases. Between these trading times, the shadow price needs to be chosen so that it is indeed optimal not to alter the portfolio at hand; compare [41]. In the present context, it is optimal to trade at all times, so that the execution price is always directly linked to the optimal strategy.

In order to calculate the dynamics of $\overline{\Delta S}^{\lambda}$ on $[\![0, \tau^{\Delta\varphi}]\!]$, we first compute the $\hat{\mathbb{Q}}$-dynamics of B^{λ}. Itô's formula and the dynamics (3.4) of $\overline{\Delta\varphi}^{\lambda}$ give

$$
\begin{aligned}
dB_t^{\lambda} &= \lambda^{-\frac{1}{p+2}}\overline{\Delta\varphi}_t^{\lambda}\,dm_t + \lambda^{-\frac{1}{p+2}}m_t\,d\overline{\Delta\varphi}_t^{\lambda} + \lambda^{-\frac{1}{p+2}}\,d\langle m, \overline{\Delta\varphi}^{\lambda}\rangle_t \\
&= \lambda^{-\frac{1}{p+2}}\overline{\Delta\varphi}_t^{\lambda}\,dm_t + \lambda^{-\frac{1}{p+2}}m_t\mu_t^{\hat{\varphi},\hat{\mathbb{Q}}}\,dt \\
&\quad - \lambda^{-\frac{2}{p+2}}\frac{1}{2}p^{-\frac{1}{p-1}}c_t^{\hat{\varphi}}m_t^2\widetilde{g}_p(B_t^{\lambda})\,dt \\
&\quad + \lambda^{-\frac{1}{p+2}}m_t\sqrt{c_t^{\hat{\varphi}}}\,dW_t^{\hat{\varphi},\hat{\mathbb{Q}}} + \lambda^{-\frac{1}{p+2}}\sqrt{c_t^{\hat{\varphi}}}\,d\langle m, W^{\hat{\varphi},\hat{\mathbb{Q}}}\rangle_t.
\end{aligned}
\tag{5.5}
$$

Now Itô's formula yields the following $\hat{\mathbb{Q}}$-dynamics for $\overline{\Delta S}^{\lambda}$ on $[\![0, \tau^{\Delta\varphi}]\!]$:

$$
\begin{aligned}
d\overline{\Delta S}_t^{\lambda} &= \lambda^{\frac{3}{p+2}}g_p(B_t^{\lambda})\,dA_t + \lambda^{\frac{3}{p+2}}A_tg_p'(B_t^{\lambda})\,dB_t^{\lambda} \\
&\quad + \frac{1}{2}\lambda^{\frac{3}{p+2}}A_tg_p''(B_t^{\lambda})\,d\langle B^{\lambda}\rangle_t + \lambda^{\frac{3}{p+2}}g_p'(B_t^{\lambda})\,d\langle A, B^{\lambda}\rangle_t \\
&= \lambda^{\frac{3}{p+2}}g_p(B_t^{\lambda})\,dA_t + \lambda^{\frac{2}{p+2}}A_tg_p'(B_t^{\lambda})\overline{\Delta\varphi}_t^{\lambda}\,dm_t \\
&\quad + \lambda^{\frac{2}{p+2}}A_tg_p'(B_t^{\lambda})m_t\mu_t^{\hat{\varphi},\hat{\mathbb{Q}}}\,dt \\
&\quad - \frac{1}{2}\lambda^{\frac{1}{p+2}}A_tc_t^{\hat{\varphi}}m_t^2p^{-\frac{1}{p-1}}\widetilde{g}_p(B_t^{\lambda})g_p'(B_t^{\lambda})\,dt \\
&\quad + \lambda^{\frac{2}{p+2}}A_tg_p'(B_t^{\lambda})m_t\sqrt{c_t^{\hat{\varphi}}}\,dW_t^{\hat{\varphi},\hat{\mathbb{Q}}} \\
&\quad + \lambda^{\frac{2}{p+2}}A_tg_p'(B_t^{\lambda})\sqrt{c_t^{\hat{\varphi}}}\,d\langle m, W^{\hat{\varphi},\hat{\mathbb{Q}}}\rangle_t + \frac{1}{2}\lambda^{\frac{1}{p+2}}A_tc^{\hat{\varphi}}m_t^2g_p''(B_t^{\lambda})\,dt \\
&\quad + \lambda^{\frac{2}{p+2}}g_p'(B_t^{\lambda})m_t\sqrt{c_t^{\hat{\varphi}}}\,d\langle A, W^{\hat{\varphi},\hat{\mathbb{Q}}}\rangle_t.
\end{aligned}
\tag{5.6}
$$

In order to simplify (5.6), observe that the ODE (3.1) implies

$$
g_p''(z) = p^{-\frac{1}{p-1}}\widetilde{g}_p(z)g_p'(z) - 2z.
$$

Using this identity and reordering terms in increasing powers of $\lambda^{\frac{1}{p+2}}$, we obtain

$$
\begin{aligned}
d\overline{\Delta S}_t^{\lambda} &= -\lambda^{\frac{1}{p+2}}A_tc^{\hat{\varphi}}m_t^2B_t^{\lambda}\,dt + \lambda^{\frac{2}{p+2}}A_tg_p'(B_t^{\lambda})\overline{\Delta\varphi}_t^{\lambda}\,dm_t \\
&\quad + \lambda^{\frac{2}{p+2}}A_tm_tg_p'(B_t^{\lambda})\sqrt{c_t^{\hat{\varphi}}}\,dW_t^{\hat{\varphi},\hat{\mathbb{Q}}} \\
&\quad + \lambda^{\frac{2}{p+2}}A_tg_p'(B_t^{\lambda})m_t\mu_t^{\hat{\varphi},\hat{\mathbb{Q}}}\,dt + \lambda^{\frac{2}{p+2}}A_tg_p'(B_t^{\lambda})\sqrt{c_t^{\hat{\varphi}}}\,d\langle m, W^{\hat{\varphi},\hat{\mathbb{Q}}}\rangle_t \\
&\quad + \lambda^{\frac{2}{p+2}}g_p'(B_t^{\lambda})m_t\sqrt{c_t^{\hat{\varphi}}}\,d\langle A, W^{\hat{\varphi},\hat{\mathbb{Q}}}\rangle_t + \lambda^{\frac{3}{p+2}}g_p(B_t^{\lambda})\,dA_t.
\end{aligned}
\tag{5.7}
$$

For future reference, we note that using the definitions of m in (3.7) and of A and B^{λ} in (5.4), the first (leading-order) term can be rewritten as

$$
-\lambda^{\frac{1}{p+2}}A_tc^{\hat{\varphi}}m_t^2B_t^{\lambda} = -2^{-(p-1)}\Lambda_t(c_t^{\hat{\varphi}})^pm_t^{p+2}\overline{\Delta\varphi}_t^{\lambda} = -\gamma c_t^S\overline{\Delta\varphi}_t^{\lambda}.
\tag{5.8}
$$

Modified execution price. As explained above, we need to modify the "naive" execution price $\overline{S}^\lambda$ in two directions. First, we need to stop. To this end, we introduce the following auxiliary stopping times:

$$\tau^{\lambda,1} := \inf\{t \in [0,T] : |B_t^\lambda| > \lambda^{-\kappa_5}\} \wedge \tau^{\Delta\varphi}, \tag{5.9}$$

$$\tau^{\lambda,2} := \inf\left\{t \in [0,T] : \left|\frac{\mu_t^{\overline{S}^\lambda,\hat{\mathbb{Q}}}\sqrt{c_t^S}}{c_t^{\overline{S}^\lambda,S}}\right| > \lambda^{\kappa_6}\right\} \wedge \tau^{\lambda,1}, \tag{5.10}$$

$$\tau^{\lambda,3} := \inf\left\{t \in [0,T] : \left|\mathcal{E}\left(\int_0^\cdot \frac{\mu_t^{\overline{S}^\lambda,\hat{\mathbb{Q}}}\sqrt{c_t^S}}{c_t^{\overline{S}^\lambda,S}}\, dW^{S,\hat{\mathbb{Q}}}\right)_t - 1\right| > \lambda^{\kappa_6}\right\} \wedge \tau^{\lambda,2}, \tag{5.11}$$

$$\tau^{\lambda,4} := \inf\{t \in [0,T] : \lambda^{-\eta}|\overline{\Delta S}_t^\lambda| > \lambda^{\kappa_7}\} \wedge \tau^{\lambda,1}, \tag{5.12}$$

$$\tau^{\lambda,5} := \inf\left\{t \in [0,T] : \frac{1}{c_t^S}\left|c_t^{\overline{\Delta S}^\lambda} + 2c_t^{S,\overline{\Delta S}^\lambda}\right| > \lambda^{-\kappa_8}\right\} \wedge \tau^{\lambda,1}, \tag{5.13}$$

$$\tau^{\lambda,6} := \inf\{t \in [0,T] : \Lambda_t^{\frac{1}{p-1}} > \lambda^{-\kappa_9}\}, \tag{5.14}$$

$$\tau^{\lambda,7} := \inf\left\{t \in [0,T] : \left|\frac{\mu_t^{\overline{S}^\lambda,\hat{\mathbb{Q}}}\sqrt{c_t^S}}{c_t^{\overline{S}^\lambda,S}} + \gamma\sqrt{c_t^S}\Delta\varphi_t^\lambda\right| > \lambda^{\kappa_{10}}\right\} \wedge \tau^{\lambda,1}. \tag{5.15}$$

These are in turn used to define the following dual stopping time:

$$\tau^{\lambda,\mathrm{dual}} = \tau^{\lambda,1} \wedge \tau^{\lambda,2} \wedge \tau^{\lambda,3} \wedge \tau^{\lambda,4} \wedge \tau^{\lambda,5} \wedge \tau^{\lambda,6} \wedge \tau^{\lambda,7}. \tag{5.16}$$

Here, the constants κ_i are chosen as follows (by the definition of κ_1, κ_2 and κ_4 in (3.9) and of η in (3.8), these intervals are not empty):

$$\kappa_5 \in \left(\frac{1}{p+2} - \kappa_1, \frac{1}{3}\frac{1}{p+2}\right), \qquad \kappa_6 \in \left(\frac{2}{3}\frac{1}{p+2}, \frac{1}{p+2} - \kappa_5\right), \tag{5.17}$$

$$\begin{aligned}
&\kappa_7 \in \left(\left(\frac{2}{p+2} - \eta + \kappa_4\right) \vee \left(\frac{1}{p+2} - \frac{\eta}{2}\right), \frac{7}{3}\frac{1}{p+2} - \eta\right),\\
&\kappa_8 \in \left(0, \frac{2}{p+2} - 2\kappa_5\right),
\end{aligned} \tag{5.18}$$

$$\kappa_9 \in \left(0, \eta - \frac{2}{p+2} + \kappa_3\right), \qquad \kappa_{10} \in \left(\frac{2}{p+2} - \kappa_1, \frac{2}{p+2} - 2\kappa_5\right). \tag{5.19}$$

With these preparations, define the modified execution price as

$$S_t^\lambda = S_t + \Delta S_t^\lambda, \tag{5.20}$$

where

$$\Delta S_t^\lambda = \int_0^t \mathbb{1}_{\{s < \tau^{\lambda,\mathrm{dual}}\}}\, d\overline{\Delta S}_s^\lambda - \int_0^t \mathbb{1}_{\{\tau^{\lambda,\mathrm{dual}} \le s \le \tau^{\lambda,\mathrm{dual}} + \lambda^\eta\}} \lambda^{-\eta} \overline{\Delta S}_{\tau^{\lambda,\mathrm{dual}}}^\lambda\, ds. \tag{5.21}$$

The modified execution price S^λ coincides with the "naive" execution price $\overline{S^\lambda}$ on $[\![0, \tau^{\lambda,\mathrm{dual}}]\!]$, on $]\!]\tau^{\lambda,\mathrm{dual}}, 0, \tau^{\lambda,\mathrm{dual}} + \lambda^\eta]\!]$, S^λ is "brought back" to the frictionless price of the risky asset S via the constant drift rate $\lambda^{-\eta}\overline{\Delta S}_{\tau^{\lambda,\mathrm{dual}}}^\lambda$ (this is the other modification of the "naive" execution price), and on $]\!]\tau^{\lambda,\mathrm{dual}} + \lambda^\eta, T]\!]$, S^λ coincides with S.

Dual martingale measure. To obtain an upper duality bound in the spirit of [30], we proceed to define an equivalent measure $\hat{\mathbb{Q}}^\lambda$ that turns the modified execution price S^λ into a martingale. To this end, set

$$\zeta_t^\lambda = \frac{\mu_t^{\overline{S}^\lambda,\hat{\mathbb{Q}}}\sqrt{c_t^S}}{c_t^{\overline{S}^\lambda,S}}\mathbb{1}_{[0\le t<\tau^{\lambda,\mathrm{dual}})} - \lambda^{-\eta}\frac{\overline{\Delta S}^\lambda_{\tau^{\lambda,\mathrm{dual}}}}{\sqrt{c_t^S}}\mathbb{1}_{[\tau^{\lambda,\mathrm{dual}}\le t\le\tau^{\lambda,\mathrm{dual}}+\lambda^\eta]}, \tag{5.22}$$

define the $\hat{\mathbb{Q}}$-martingale $\hat{Z}^\lambda$ by[24]

$$\hat{Z}_t^\lambda := \mathcal{E}\Big(-\int_0^\cdot \zeta_s^\lambda\,dW_s^{S,\hat{\mathbb{Q}}}\Big)_t, \tag{5.23}$$

and define the measure $\hat{\mathbb{Q}}^\lambda \approx \hat{\mathbb{Q}}$ on $\mathcal{F}_T$ by

$$\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}} = \hat{Z}_T^\lambda. \tag{5.24}$$

Girsanov's theorem ensures that S^λ is a local $\hat{\mathbb{Q}}^\lambda$-martingale.

To ease the notation in upcoming estimates, we also define

$$\Delta\hat{Z}_t^\lambda := \hat{Z}_t^\lambda - 1. \tag{5.25}$$

Note that $\Delta\hat{Z}^\lambda$ is as $\hat{Z}^\lambda$ a $\hat{\mathbb{Q}}^\lambda$-martingale.

We proceed to show that the wealth process of any admissible strategy is a $\hat{\mathbb{Q}}^\lambda$-martingale when evaluated with the modified execution price S^λ.

LEMMA 5.2. *Suppose that $E^{\hat{\mathbb{Q}}}[\exp(\varepsilon((c^S)^{-1})^*_T)] < \infty$ (which is part of Assumptions 3). Then the stochastic integral $\int_0^\cdot \varphi_t\,dS_t^\lambda$ is a square-integrable $\hat{\mathbb{Q}}^\lambda$-martingale for any admissible strategy $\varphi \in \Phi^\lambda$.*

PROOF. Since ΔS^λ does not accumulate quadratic variation on $]\!]\tau^{\lambda,\mathrm{dual}}, T]\!]$, the instantaneous quadratic variation of S^λ satisfies

$$c_t^{S^\lambda} = c_t^{S+\Delta S^\lambda} = c_t^S + c_t^{\Delta S^\lambda} + 2c_t^{S,\Delta S^\lambda} \le c_t^S(1+\lambda^{-\kappa_8}),\quad t\in[0,T].$$

Here, the last inequality follows from (5.13) (noting that ΔS^λ and $\overline{\Delta S}^\lambda$ are indistinguishable on $[\![0,\tau^{\lambda,\mathrm{dual}}]\!]$). Together with Hölder's inequality, Lemma 5.3 below, and the admissibility of φ, this yields

$$\begin{aligned}
&\mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\Big[\int_0^T \varphi_t^2\,d\langle S^\lambda\rangle_t\Big]\\
&\quad\le (1+\lambda^{-\kappa_8})\mathbb{E}_{\hat{\mathbb{Q}}}\Big[\hat{Z}_T^\lambda\int_0^T \varphi_t^2 c_t^S\,dt\Big]\\
&\quad\le (1+\lambda^{-\kappa_8})\mathbb{E}_{\hat{\mathbb{Q}}}[(\hat{Z}_T^\lambda)^{\frac{1+a}{a}}]^{\frac{a}{1+a}}\mathbb{E}_{\hat{\mathbb{Q}}}\Big[\Big(\int_0^T \varphi_t^2 c_t^S\,dt\Big)^{1+a}\Big]^{\frac{1}{1+a}} < \infty,
\end{aligned}$$

for some $0 < a \le \varepsilon$. Therefore, $\varphi \in L^2_{\hat{\mathbb{Q}}^\lambda}(S^\lambda)$ and the assertion follows. □

[24] It is shown in Lemma 5.3 below that the local $\hat{\mathbb{Q}}$-martingale $\hat{Z}^\lambda$ is indeed a $\hat{\mathbb{Q}}$-martingale.

Convex duality estimates. To derive the dual upper bound in Proposition 5.1, we follow [30] in using the convex conjugates of both the utility function U and the instantaneous trading cost $x \mapsto \Psi_t(x) = \lambda_t |x|^p$:[25]

$$\begin{aligned} \widetilde{\Psi}_t(y) &= \sup_{x\in\mathbb{R}}\{xy - \Psi_t(x)\} = \frac{p-1}{p(\lambda_t p)^{\frac{1}{p-1}}}|y|^{\frac{p}{p-1}}, \\ \widetilde{\Psi}_t'(y) &= \frac{1}{(\lambda_t p)^{\frac{1}{p-1}}}|y|^{\frac{1}{p-1}}\operatorname{sgn}(y), \end{aligned} \tag{5.26}$$

$$\widetilde{U}(y) = \sup_{x\in\mathbb{R}}\{U(x) - xy\} = \frac{y}{\gamma}\left(\log\left(\frac{y}{\gamma}\right) - 1\right), \tag{5.27}$$

$$\widetilde{U}'(y) = \frac{1}{\gamma}\log\left(\frac{y}{\gamma}\right),$$

$$\widetilde{U}''(y) = \frac{1}{\gamma y}, \qquad \widetilde{U}'''(y) = -\frac{1}{\gamma y^2}. \tag{5.28}$$

Let φ be any admissible trading strategy. Since the terminal risky position is zero, two integrations by parts, the definition of $\widetilde{\Psi}$, and $S_0^\lambda = S_0$ yield

$$\begin{aligned} X_T^\varphi &= x + \int_0^T \varphi_t\,dS_t - \int_0^T \lambda_t|\dot\varphi_t|^p\,dt \\ &= x - \varphi_0 S_0 - \int_0^T \dot\varphi_t S_t\,dt - \int_0^T \lambda_t|\dot\varphi_t|^p\,dt \\ &= x - \varphi_0 S_0^\lambda - \int_0^T \dot\varphi_t S_t^\lambda\,dt + \int_0^T \dot\varphi_t(S_t^\lambda - S_t)\,dt - \int_0^T \lambda_t|\dot\varphi_t|^p\,dt \\ &\le x + \int_0^T \varphi_t\,dS_t^\lambda + \int_0^T \widetilde{\Psi}_t(\Delta S_t^\lambda)\,dt. \end{aligned} \tag{5.29}$$

PROOF OF PROPOSITION 5.1. By definition of the convex conjugate $\widetilde{U}(y) = \sup_{x\in\mathbb{R}}\{U(x) - xy\}$, we have the pointwise inequality

$$U(X_T^\varphi) \le \widetilde{U}\left(\hat y\frac{d\hat{\mathbb{Q}}^\lambda}{d\mathbb{P}}\right) + \hat y\frac{d\hat{\mathbb{Q}}^\lambda}{d\mathbb{P}}X_T^\varphi = \widetilde{U}(\hat y\hat Z_T\hat Z_T^\lambda) + \hat y\hat Z_T\hat Z_T^\lambda X_T^\varphi, \tag{5.30}$$

where $d\hat Z_T := d\hat{\mathbb{Q}}/d\mathbb{P}$ denotes the Radon–Nikodým density of the frictionless minimal entropy martingale measure. We proceed to establish an upper bound for $\widetilde{U}(\hat y\hat Z_T\hat Z_T^\lambda)$. For fixed ω, a second-order Taylor expansion with Lagrange remainder term gives

$$\begin{aligned} \widetilde{U}(\hat y\hat Z_T\hat Z_T^\lambda) &= \widetilde{U}(\hat y\hat Z_T) + \widetilde{U}'(\hat y\hat Z_T)\hat y\hat Z_T\Delta\hat Z_T^\lambda \\ &\quad + \frac{1}{2}\widetilde{U}''(\hat y\hat Z_T)\hat y^2\hat Z_T^2(\Delta\hat Z_T^\lambda)^2 \\ &\quad + \frac{1}{6}\widetilde{U}'''(\xi(\omega)\hat y\hat Z_T)\hat y^3\hat Z_T^3(\Delta\hat Z_T^\lambda)^3, \end{aligned} \tag{5.31}$$

where $\xi(\omega)$ takes values in the interval with endpoints 1 and $\hat Z_T^\lambda$. Using $-\widetilde{U}' = (U')^{-1}$ together with the first-order condition $U'(x + \int_0^T \hat\varphi_t\,dS_t) = \hat y\hat Z_T$, the explicit expressions in

[25] The importance of the dual friction $\widetilde{\Psi}$ was first recognized in [21], where it is used to establish a superhedging theorem.

(5.28) for the derivatives $\widetilde{U}''$ and $\widetilde{U}'''$ and the facts that $\widetilde{U}'''(\xi(\omega)\hat{y}\hat{Z}_T)(\Delta\hat{Z}_T^\lambda)^3 \le 0$ if $\hat{Z}_T^\lambda \ge 1$, and $\xi(\omega)^{-2} \le (\hat{Z}_T^\lambda)^{-2}$ for $\hat{Z}_T^\lambda \le 1$ and plugging all this into (5.31) yields

$$\begin{aligned}\widetilde{U}(\hat{y}\hat{Z}_T\hat{Z}_T^\lambda) \le \widetilde{U}(\hat{y}\hat{Z}_T) - \hat{y}\hat{Z}_T\Delta\hat{Z}_T^\lambda\Big(x + \int_0^T \hat{\varphi}_t\,dS_t\Big)& \\ + \frac{1}{2\gamma}\hat{y}\hat{Z}_T(\Delta\hat{Z}_T^\lambda)^2 + \frac{1}{6\gamma}\frac{\hat{y}\hat{Z}_T}{(\hat{Z}_T^\lambda)^2}|\Delta\hat{Z}_T^\lambda|^3.&\end{aligned} \tag{5.32}$$

Combining (5.32) with (5.30) and using $\widetilde{U}(\hat{y}\hat{Z}_T) = U(x + \int_0^T \hat{\varphi}_t\,dS_t) - \hat{y}\hat{Z}_T(x + \int_0^T \hat{\varphi}_t\,dS_t)$ and (5.29), we in turn obtain

$$\begin{aligned}&U(X_T^\varphi) - U\Big(x + \int_0^T \hat{\varphi}_t\,dS_t\Big) \\ &\quad\le \hat{y}\hat{Z}_T\hat{Z}_T^\lambda\Big(X_T^\varphi - x - \int_0^T \hat{\varphi}_t\,dS_t\Big) \\ &\qquad + \frac{1}{2\gamma}\hat{y}\hat{Z}_T(\Delta\hat{Z}_T^\lambda)^2 + \frac{1}{6\gamma}\hat{y}\hat{Z}_T\frac{|\Delta\hat{Z}_T^\lambda|^3}{(\hat{Z}_T^\lambda)^2} \\ &\quad\le \hat{y}\hat{Z}_T\hat{Z}_T^\lambda\Big(\int_0^T \varphi_t\,dS_t^\lambda\Big) + \hat{y}\hat{Z}_T\hat{Z}_T^\lambda\Big(\int_0^T \widetilde{\Psi}_t(\Delta S_t^\lambda)\,dt\Big) \\ &\qquad - \hat{y}\hat{Z}_T\hat{Z}_T^\lambda\Big(\int_0^T \hat{\varphi}_t\,dS_t\Big) + \frac{1}{2\gamma}\hat{y}\hat{Z}_T(\Delta\hat{Z}_T^\lambda)^2 + \frac{1}{6\gamma}\hat{y}\hat{Z}_T\frac{|\Delta\hat{Z}_T^\lambda|^3}{(\hat{Z}_T^\lambda)^2}.\end{aligned} \tag{5.33}$$

Now take $\mathbb{P}$-expectations in (5.33). We consider each of the five terms on the right-hand side of (5.33) separately.

For the first term, we use Bayes' theorem and the fact that $\int_0^\cdot \varphi_t\,dS_t^\lambda$ is a $\hat{\mathbb{Q}}^\lambda$-martingale by Lemma 5.2. This gives

$$\mathbb{E}\Big[\hat{y}\hat{Z}_T\hat{Z}_T^\lambda\Big(\int_0^T \varphi_t\,dS_t^\lambda\Big)\Big] = \hat{y}\mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\Big[\int_0^T \varphi_t\,dS_t^\lambda\Big] = 0. \tag{5.34}$$

For the second term, we use Bayes' theorem and Lemma 5.4 to obtain

$$\begin{aligned}&\mathbb{E}\Big[\hat{y}\hat{Z}_T Z_T^\lambda\Big(\int_0^T \widetilde{\Psi}_t(\Delta S_t^\lambda)\,dt\Big)\Big] \\ &\quad= \hat{y}\mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\Big[\int_0^T \widetilde{\Psi}_t(\Delta S_t^\lambda)\,dt\Big] \\ &\quad= \hat{y}(p-1)\mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_0^{\tau^{\Delta\varphi}} \lambda_t|\dot{\varphi}_t^\lambda|^p\,dt\Big] + o(\lambda^{\frac{2}{p+2}}).\end{aligned} \tag{5.35}$$

For the third term, we apply Bayes' theorem, the decomposition $\hat{\varphi} = \varphi^\lambda + \Delta\varphi^\lambda$ and Lemma 5.5 and 5.7. This gives

$$\begin{aligned}&\mathbb{E}\Big[-\hat{y}\hat{Z}_T\hat{Z}_T^\lambda\Big(\int_0^T \hat{\varphi}_t\,dS_t\Big)\Big] \\ &\quad= -\hat{y}\Big(\mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\Big[\int_0^T \varphi_t^\lambda\,dS_t\Big] + \mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\Big[\int_0^T \Delta\varphi_t^\lambda\,dS_t\Big]\Big) \\ &\quad= -\hat{y}p\mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_0^{\tau^{\Delta\varphi}} \lambda_t|\dot{\varphi}_t^\lambda|^p\,dt\Big] - \hat{y}\gamma\mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_0^{\tau^{\Delta\varphi}} c_t^S(\Delta\varphi_t^\lambda)^2\,dt\Big] \\ &\qquad + o(\lambda^{\frac{2}{p+2}}).\end{aligned} \tag{5.36}$$

For the fourth term, Bayes' theorem and Lemma 5.8 yield

$$\mathbb{E}\bigg[\frac{1}{2\gamma}\hat{y}\hat{Z}_T(\Delta\hat{Z}^\lambda_T)^2\bigg] = \frac{\hat{y}}{2\gamma}\mathbb{E}_{\hat{\mathbb{Q}}}[(\Delta\hat{Z}^\lambda_T)^2]$$
$$(5.37)\qquad = \frac{\hat{y}\gamma}{2}\mathbb{E}_{\hat{\mathbb{Q}}}\bigg[\int_0^{\tau^{\Delta\varphi}} c^S_t(\Delta\varphi^\lambda_t)^2\,dt\bigg] + o(\lambda^{\frac{2}{p+2}}).$$

Finally, for the last term, we use Bayes' theorem and Lemma 5.9 to obtain

$$(5.38)\qquad \mathbb{E}\bigg[\frac{1}{6\gamma}\hat{y}\hat{Z}_T\frac{|\Delta\hat{Z}^\lambda_T|^3}{(\hat{Z}^\lambda_T)^2}\bigg] = \frac{\hat{y}}{6\gamma}\mathbb{E}_{\hat{\mathbb{Q}}}\bigg[\frac{|\Delta\hat{Z}^\lambda_T|^3}{(\hat{Z}^\lambda_T)^2}\bigg] = o(\lambda^{\frac{2}{p+2}}).$$

Combining (5.33)–(5.38) now yields the claimed upper duality bound from Proposition 5.1. □

5.3. *Auxiliary estimates.* For better readability, this section collects auxiliary estimates used in the proof of Proposition 5.1. We first establish a maximal inequality for the density process $\hat{Z}^\lambda$ defined in (5.23). This estimate is used both in the proof of Lemma 5.2 and in several other auxiliary results below.

LEMMA 5.3. *Let $k \in \mathbb{R}$. Suppose that $E^{\hat{\mathbb{Q}}}[\exp(\varepsilon((c^S)^{-1})^*_T)] < \infty$ (which is part of Assumptions 3). Then there exist constants C_k and $\lambda_k > 0$ such that, for all $\lambda \in (0, \lambda_k)$,*

$$\mathbb{E}_{\hat{\mathbb{Q}}}[((\hat{Z}^\lambda)^k)^*_T] \le C_k.$$

PROOF. To prove the assertion, we first establish three auxiliary claims.

CLAIM 1. *For any $\alpha \in \mathbb{R}_+$, there are constants $c_\alpha, \bar{\lambda}_\alpha > 0$ such that, for all $\lambda \in (0, \bar{\lambda}_\alpha)$:*

$$\mathbb{E}_{\hat{\mathbb{Q}}}\bigg[\bigg(\exp\bigg(\alpha\int_0^\cdot(\zeta^\lambda_s)^2\,ds\bigg)\bigg)^*_T\bigg] \le c_\alpha,$$

for some positive constants $\bar{\lambda}_\alpha$ and c_α depending only on α.

PROOF OF CLAIM 1. For $\alpha \in \mathbb{R}_+$, the integrability assumption on c^S shows that there exists $\bar{\lambda}_\alpha > 0$ such that

$$c_\alpha := \mathbb{E}_{\hat{\mathbb{Q}}}[\exp(\alpha T(\bar{\lambda}_\alpha)^{2\kappa_6} + \alpha(\bar{\lambda}_\alpha)^{\eta+2\kappa_7}((c^S)^{-1})^*_T)] < \infty.$$

Since $(\zeta^\lambda_t)^2$ is nonnegative, the definition (5.22) of the process ζ^λ, the definitions of the stopping times $\tau^{\lambda,2}$, $\tau^{\lambda,3}$ and $\tau^{\lambda,4}$ in (5.10–5.12), and the choice of c_α show that, for $\lambda \in (0, \bar{\lambda}_\alpha)$:

$$\mathbb{E}_{\hat{\mathbb{Q}}}\bigg[\bigg(\exp\bigg(\alpha\int_0^\cdot(\zeta^\lambda_s)^2\,ds\bigg)\bigg)^*_T\bigg]$$
$$\le \mathbb{E}_{\hat{\mathbb{Q}}}\bigg[\exp\bigg(\alpha\int_0^T(\zeta^\lambda_s)^2\,ds\bigg)\bigg]$$
$$(5.39)\qquad = \mathbb{E}_{\hat{\mathbb{Q}}}\bigg[\exp\bigg(\alpha\int_0^{\tau^{\lambda,\mathrm{dual}}}\bigg(\frac{\mu^{\overline{S}^\lambda,\hat{\mathbb{Q}}}_s\sqrt{c^S_s}}{c^{\overline{S}^\lambda,S}_s}\bigg)^2\,ds$$
$$+ \alpha\int_{\tau^{\lambda,\mathrm{dual}}}^{\tau^{\lambda,\mathrm{dual}}+\lambda^\eta}\bigg(\lambda^{-\eta}\frac{\overline{\Delta S}^\lambda_{\tau^{\lambda,\mathrm{dual}}}}{\sqrt{c^S_s}}\bigg)^2\,ds\bigg)\bigg]$$
$$\le \mathbb{E}_{\hat{\mathbb{Q}}}[\exp(\alpha T\lambda^{2\kappa_6} + \alpha\lambda^{\eta+2\kappa_7}((c^S)^{-1})^*_T)] \le c_\alpha.$$

□

CLAIM 2. *For any* $\alpha \in \mathbb{R}$ *and* $\lambda \in (0, \bar{\lambda}_{\alpha^2/2})$, *the process* $\mathcal{E}(-\alpha \int_0^\cdot \zeta_s^\lambda \, dW_s^{S,\hat{\mathbb{Q}}})$ *is a* $\hat{\mathbb{Q}}$-*martingale.*

PROOF OF CLAIM 2. This follows from Claim 1 and Novikov's criterion. □

CLAIM 3. *For any* $\alpha \in \mathbb{R}$ *and* $\lambda \in (0, \bar{\lambda}_{6\alpha^2} \wedge \bar{\lambda}_{8\alpha^2})$,

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\mathcal{E}\left(-\alpha \int_0^\cdot \zeta_s^\lambda \, dW_s^{S,\hat{\mathbb{Q}}}\right)_T^2\right] \leq c_{6\alpha^2}.$$

PROOF OF CLAIM 3. By the Cauchy–Schwarz inequality, the fact that $\mathcal{E}(-4\alpha \times \int_0^\cdot \zeta_s^\lambda \, dW_s^{S,\hat{\mathbb{Q}}})$ is a supermartingale and Claim 1, we obtain

$$\begin{aligned}
&\mathbb{E}_{\hat{\mathbb{Q}}}\left[\mathcal{E}\left(-\alpha \int_0^\cdot \zeta_s^\lambda \, dW_s^{S,\hat{\mathbb{Q}}}\right)_T^2\right] \\
&\quad = \mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(-2\alpha \int_0^T \zeta_t^\lambda \, dW_t^{S,\hat{\mathbb{Q}}} - \frac{1}{2}\int_0^T (8\alpha^2 - 6\alpha^2)(\zeta_t^\lambda)^2 \, dt\right)\right] \\
&\quad \leq \mathbb{E}_{\hat{\mathbb{Q}}}\left[\mathcal{E}\left(-4\alpha \int_0^\cdot \zeta_s^\lambda \, dW_s^{S,\hat{\mathbb{Q}}}\right)_T\right]^{\frac{1}{2}} \mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(6\alpha^2 \int_0^T (\zeta_t^\lambda)^2 \, dt\right)\right]^{\frac{1}{2}} \leq c_{6\alpha^2}
\end{aligned}$$

as asserted. □

To complete the proof of Lemma 5.3, now let $k \in \mathbb{R}$ and set $\lambda_k := \bar{\lambda}_{k^2/2} \wedge \bar{\lambda}_{|k^2-k|} \wedge \bar{\lambda}_{6k^2} \wedge \bar{\lambda}_{8\alpha^2}$. Then for $\lambda \in (0, \lambda_k)$,the Cauchy–Schwarz inequality, Claim 2, Doob's inequality and Claims 1 and 3 yield the asserted estimate:

$$\begin{aligned}
&\mathbb{E}[((\hat{Z}^\lambda)^k)_T^*] \\
&\quad = \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\exp\left(-\int_0^\cdot k\zeta_s^\lambda \, dW_s^{S,\hat{\mathbb{Q}}} - \frac{1}{2}\int_0^\cdot k^2(\zeta_s^\lambda)^2 \, ds \right.\right.\right. \\
&\quad\quad \left.\left.\left. + \frac{1}{2}\int_0^\cdot (k^2-k)(\zeta_s^\lambda)^2 \, ds\right)\right)_T^*\right] \\
&\quad \leq \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\mathcal{E}\left(-k\int_0^\cdot \zeta_s^\lambda \, dW_s^{S,\hat{\mathbb{Q}}}\right)^2\right)_T^*\right]^{\frac{1}{2}} \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\exp\left((|k^2-k|)\int_0^t (\zeta_s^\lambda)^2 \, ds\right)\right)_T^*\right]^{\frac{1}{2}} \\
&\quad \leq 2\mathbb{E}_{\hat{\mathbb{Q}}}\left[\mathcal{E}\left(-k\int_0^\cdot \zeta_s^\lambda \, dW_s^{S,\hat{\mathbb{Q}}}\right)_T^2\right]^{\frac{1}{2}} (c_{|k^2-k|})^{\frac{1}{2}} \leq 2(c_{6k^2})^{\frac{1}{2}}(c_{|k^2-k|})^{\frac{1}{2}} =: C_k.
\end{aligned}$$

□

Next, we provide a series of estimates that are used in the proof of Proposition 5.1.

LEMMA 5.4. *Suppose Assumptions 1, 2 and 3 are satisfied. Then*

$$\mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_0^T \widetilde{\Psi}_t(\Delta S_t^\lambda) \, dt\right] = (p-1)\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}} \lambda_t |\dot{\varphi}_t^\lambda|^p \, dt\right] + o(\lambda^{\frac{2}{p+2}}).$$

PROOF. Using the definitions of $\widetilde{\Psi}_t$ in (5.26) and ΔS_t^λ in (5.21), we obtain

$$
\begin{aligned}
&\mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_0^T \widetilde{\Psi}_t(\Delta S_t^\lambda)\,dt\right]\\
&\quad = (p-1)\mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}} \lambda_t|\dot{\varphi}_t^\lambda|^p\,dt\right]\\
&\qquad + (p-1)\mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_{\tau^{\lambda,\mathrm{dual}}}^{\tau^{\lambda,\mathrm{dual}}+\lambda^\eta} \bigl(1-\lambda^{-\eta}(t-\tau^{\lambda,\mathrm{dual}})\bigr)^{\frac{p}{p-1}}\left(\frac{\Lambda_{\tau^{\lambda,\mathrm{dual}}}}{\Lambda_t}\right)^{\frac{1}{p-1}}\right.\\
&\qquad \left.\times\,\lambda_{\tau^{\lambda,\mathrm{dual}}}|\dot{\varphi}^\lambda_{\tau^{\lambda,\mathrm{dual}}}|^p\,dt\right].
\end{aligned}
\tag{5.40}
$$

By Lemma 5.6 below, it suffices to show that the second term on the right-hand side of (5.40) is of order $o(\lambda^{\frac{2}{p+2}})$. To this end, we use that $\lambda_{\tau^{\lambda,\mathrm{dual}}}|\dot{\varphi}^\lambda_{\tau^{\lambda,\mathrm{dual}}}|^p \le \lambda^{\kappa_3}$ since $\tau^{\lambda,\mathrm{dual}} \le \tau^{\Delta\varphi}$ (see the definition of $\tau^{\Delta\varphi}$ in (3.10)). Using this, the definition of $\tau^{\lambda,6}$ in (5.14), Bayes' theorem, Hölder's inequality, the integrability condition on Λ^{-1} from Assumption 3 and Lemma 5.3, we obtain

$$
\begin{aligned}
&\mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_{\tau^{\lambda,\mathrm{dual}}}^{\tau^{\lambda,\mathrm{dual}}+\lambda^\eta} \bigl(1-\lambda^{-\eta}(t-\tau^{\lambda,\mathrm{dual}})\bigr)^{\frac{p}{p-1}}\left(\frac{\Lambda_{\tau^{\lambda,\mathrm{dual}}}}{\Lambda_t}\right)^{\frac{1}{p-1}}\lambda_{\tau^{\lambda,\mathrm{dual}}}|\dot{\varphi}^\lambda_{\tau^{\lambda,\mathrm{dual}}}|^p\,dt\right]\\
&\quad \le \lambda^{\eta+\kappa_3-\kappa_9}\mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\bigl[(\Lambda^{-\frac{1}{p-1}})_T^*\bigr] \le \lambda^{\eta+\kappa_3-\kappa_9}\mathbb{E}_{\hat{\mathbb{Q}}}\bigl[(\Lambda^{-\frac{1+\varepsilon}{p-1}})_T^*\bigr]^{\frac{1}{1+\varepsilon}}\mathbb{E}_{\hat{\mathbb{Q}}}\bigl[(\hat{Z}_T^\lambda)^{\frac{1+\varepsilon}{\varepsilon}}\bigr]^{\frac{\varepsilon}{1+\varepsilon}}\\
&\quad = o\bigl(\lambda^{\frac{2}{p+2}}\bigr). \qquad \square
\end{aligned}
$$

LEMMA 5.5. *Suppose Assumptions 1, 2 and 3 are satisfied. Then*

$$
\mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_0^T \varphi_t^\lambda\,dS_t\right] = p\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}} \lambda_t|\dot{\varphi}_t^\lambda|^p\,dt\right] + o\bigl(\lambda^{\frac{2}{p+2}}\bigr).
$$

PROOF. Using that $S = S^\lambda - \Delta S^\lambda$ and taking into account that $\int_0^\cdot \varphi_t^\lambda\,dS_t^\lambda$ is a $\hat{\mathbb{Q}}^\lambda$-martingale by Lemma 5.2, it suffices to show that

$$
\mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_0^T -\hat{\varphi}_t^\lambda\,d\Delta S_t^\lambda\right] = p\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}} \lambda_t|\dot{\varphi}_t^\lambda|^p\,dt\right] + o\bigl(\lambda^{\frac{2}{p+2}}\bigr).
\tag{5.41}
$$

To establish (5.41), we use that ΔS^λ vanishes on $\rrbracket\tau^{\lambda,\mathrm{dual}}+\lambda^\eta, T\rrbracket$, integrate by parts (recalling that $\Delta S^\lambda_{\tau^{\lambda,\mathrm{dual}}+\lambda^\eta} = \Delta S_0^\lambda = 0$), and recall the definition of ΔS^λ in (5.21). This gives

$$
\begin{aligned}
&\mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_0^T -\varphi_t^\lambda\,d\Delta S_t^\lambda\right]\\
&\quad = \mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}+\lambda^\eta} -\varphi_t^\lambda\,d\Delta S_t^\lambda\right] = \mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}+\lambda^\eta} \Delta S_t^\lambda\dot{\varphi}_t^\lambda\,dt\right]\\
&\quad = p\mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}} \lambda_t|\dot{\varphi}_t^\lambda|^p\,dt\right] + \mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_{\tau^{\lambda,\mathrm{dual}}}^{\tau^{\lambda,\mathrm{dual}}+\lambda^\eta} \Delta S_t^\lambda\dot{\varphi}_t^\lambda\,dt\right].
\end{aligned}
\tag{5.42}
$$

By Lemma 5.6 below, it suffices to show that the second term on the right-hand side of (5.42) is of order $o(\lambda^{\frac{2}{p+2}})$. To this end, we use the definition of ΔS^λ, φ^λ, $\overline{\Delta S}^\lambda$, $\tau^{\Delta\varphi}$ in (5.21), (3.11), (5.2), (3.10), $\lambda_t = \lambda\Lambda_t$, the definition (5.12) and (5.14) of $\tau^{\lambda,4}$ and $\tau^{\lambda,6}$, Hölder's inequality, Lemmas B.1 and 5.3, the integrability condition on Λ^{-1} from Assumption 3 and the choice of the constants κ_3, κ_4, κ_7 and κ_9, in (3.9), (5.18) and (5.19) to deduce

$$\begin{aligned}
&\left|\mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_{\tau^{\lambda,\mathrm{dual}}}^{\tau^{\lambda,\mathrm{dual}}+\lambda^\eta} \Delta S_t^\lambda \dot{\varphi}_t^\lambda \,dt\right]\right| \\
&\quad\le \mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_{\tau^{\lambda,\mathrm{dual}}}^{\tau^{\lambda,\mathrm{dual}}+\lambda^\eta} |\Delta S^\lambda_{\tau^{\lambda,\mathrm{dual}}}||\dot{\varphi}_t^\lambda|\,dt\right] \\
&\quad\le \mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_{\tau^{\lambda,\mathrm{dual}}}^{\tau^{\Delta\varphi}\wedge(\tau^{\lambda,\mathrm{dual}}+\lambda^\eta)} |\Delta S^\lambda_{\tau^{\lambda,\mathrm{dual}}}||\dot{\varphi}_t^\lambda|\,dt\right] \\
&\qquad+ \mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_{\tau^{\Delta\varphi}\wedge(\tau^{\lambda,\mathrm{dual}}+\lambda^\eta)}^{\tau^{\lambda,\mathrm{dual}}+\lambda^\eta} |\Delta S^\lambda_{\tau^{\lambda,\mathrm{dual}}}|\lambda^{-\eta}|\varphi^\lambda_{\tau^{\Delta\varphi}}|\,dt\right] \\
&\quad= \mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_{\tau^{\lambda,\mathrm{dual}}}^{\tau^{\Delta\varphi}\wedge(\tau^{\lambda,\mathrm{dual}}+\lambda^\eta)} p\lambda_{\tau^{\lambda,\mathrm{dual}}}^{\frac{p-1}{p}}|\dot{\varphi}_{\tau^{\lambda,\mathrm{dual}}}|^{p-1}\lambda_t^{\frac{1}{p}}|\dot{\varphi}_t^\lambda|\left(\frac{\Lambda_{\tau^{\lambda,\mathrm{dual}}}}{\Lambda_t}\right)^{\frac{1}{p}}\,dt\right] \\
&\qquad+ \mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_{\tau^{\Delta\varphi}\wedge(\tau^{\lambda,\mathrm{dual}}+\lambda^\eta)}^{\tau^{\lambda,\mathrm{dual}}+\lambda^\eta} |\Delta S^\lambda_{\tau^{\lambda,\mathrm{dual}}}|\lambda^{-\eta}|\varphi^\lambda_{\tau^{\Delta\varphi}}|\,dt\right] \\
&\quad\le p\lambda^{\kappa_3-\frac{p-1}{p}\kappa_9+\eta}\mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\big[(\Lambda^{-\frac{1}{p}})^*_T\big]+\lambda^{\kappa_7-\kappa_4+\eta} \\
&\quad\le p\lambda^{\kappa_3-\frac{p-1}{p}\kappa_9+\eta}\mathbb{E}_{\hat{\mathbb{Q}}}\big[(\hat{Z}_T^\lambda)^{\frac{1+\varepsilon}{\varepsilon}}\big]^{\frac{\varepsilon}{1+\varepsilon}}\mathbb{E}_{\hat{\mathbb{Q}}}\big[(\Lambda^{-\frac{1+\varepsilon}{p}})^*_T\big]^{\frac{1}{1+\varepsilon}}+\lambda^{\kappa_7-\kappa_4+\eta} = o\big(\lambda^{\frac{2}{p+2}}\big).
\end{aligned}$$

This completes the proof. □

LEMMA 5.6. *Suppose Assumptions 1, 2 and 3 are satisfied. Then*

$$\mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}}\lambda_t|\dot{\varphi}_t^\lambda|^p\,dt\right] = \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}}\lambda_t|\dot{\varphi}_t^\lambda|^p\,dt\right]+o\big(\lambda^{\frac{2}{p+2}}\big).$$

PROOF. By Bayes' theorem and since $\hat{Z}^\lambda = 1+\Delta\hat{Z}^\lambda$, we have

$$\begin{aligned}
\mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}}\lambda_t|\dot{\varphi}_t^\lambda|^p\,dt\right] &= \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}}\lambda_t|\dot{\varphi}_t^\lambda|^p\,dt\right] \\
&\quad+\mathbb{E}_{\hat{\mathbb{Q}}}\left[\Delta\hat{Z}^\lambda_{\tau^{\lambda,\mathrm{dual}}}\int_0^{\tau^{\lambda,\mathrm{dual}}}\lambda_t|\dot{\varphi}_t^\lambda|^p\,dt\right].
\end{aligned}$$

The claim now follows from dominated convergence, using that $\mathbb{E}_{\hat{\mathbb{Q}}}[\int_0^{\tau^{\Delta\varphi}}\lambda_t|\dot{\varphi}_t^\lambda|^p\,dt] = O(\lambda^{\frac{2}{p+2}})$ by Proposition 4.1, taking into account that $\tau^{\lambda,\mathrm{dual}}$ converges in probability to $\tau^{\Delta\varphi}$ as $\lambda\to 0$ by Lemma D.2, and observing that $|\Delta\hat{Z}^\lambda_{\tau^{\lambda,\mathrm{dual}}}|\le\lambda^{\kappa_6}$ by the definition of $\tau^{\lambda,3}$ in (5.11). □

LEMMA 5.7. *Suppose Assumptions 1, 2 and 3 are satisfied. Then*

$$\mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_0^T\Delta\varphi_t^\lambda\,dS_t\right] = \gamma\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}}c_t^S(\Delta\varphi_t^\lambda)^2\,dt\right]+o\big(\lambda^{\frac{2}{p+2}}\big).$$

PROOF. Bayes' theorem, the fact that $\hat{Z}^\lambda$ and $\int_0^\cdot\Delta\varphi_t^\lambda\,dS_t$ are square-integrable $\hat{\mathbb{Q}}$-martingales by Theorem 3.3 and Lemma 5.3, and the fact that $\zeta^\lambda = 0$ on $\rrbracket\tau^{\lambda,\mathrm{dual}}+\lambda^\eta, T\rrbracket$

yield

$$
\begin{aligned}
&\mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\Big[\int_0^T \Delta\varphi_t^\lambda\,dS_t\Big]\\
&\quad= \mathbb{E}_{\hat{\mathbb{Q}}}\Big[(1+\Delta\hat{Z}_T^\lambda)\Big(\int_0^T \Delta\varphi_t^\lambda\,dS_t\Big)\Big] = \mathbb{E}_{\hat{\mathbb{Q}}}\Big[\Delta\hat{Z}_T^\lambda\Big(\int_0^T \Delta\varphi_t^\lambda\,dS_t\Big)\Big]\\
&\quad= \mathbb{E}_{\hat{\mathbb{Q}}}\Big[\Big\langle \hat{Z}^\lambda, \int_0^\cdot \Delta\varphi_t^\lambda\,dS_t\Big\rangle_T\Big] = -\mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_0^T (1+\Delta\hat{Z}_t^\lambda)\zeta_t^\lambda\Delta\varphi_t^\lambda\sqrt{c_t^S}\,dt\Big]\\
&\quad= -\mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_0^{\tau^{\lambda,\mathrm{dual}}} (1+\Delta\hat{Z}_t^\lambda)\zeta_t^\lambda\Delta\varphi_t^\lambda\sqrt{c_t^S}\,dt\Big]\\
&\qquad - \mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_{\tau^{\lambda,\mathrm{dual}}}^{\tau^{\lambda,\mathrm{dual}}+\lambda^\eta} (1+\Delta\hat{Z}_t^\lambda)\zeta_t^\lambda\Delta\varphi_t^\lambda\sqrt{c_t^S}\,dt\Big]\\
&\quad= \mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_0^{\tau^{\lambda,\mathrm{dual}}} (1+\Delta\hat{Z}_t^\lambda)\gamma(\Delta\varphi_t^\lambda)^2 c_t^S\,dt\Big]\\
&\qquad - \mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_{\tau^{\lambda,\mathrm{dual}}}^{\tau^{\lambda,\mathrm{dual}}+\lambda^\eta} \hat{Z}_t^\lambda\zeta_t^\lambda\Delta\varphi_t^\lambda\sqrt{c_t^S}\,dt\Big]\\
&\qquad - \mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_0^{\tau^{\lambda,\mathrm{dual}}} (1+\Delta\hat{Z}_t^\lambda)\Delta\varphi_t^\lambda\big(\zeta_t^\lambda+\gamma\Delta\varphi_t^\lambda\sqrt{c^S}_t\big)\sqrt{c_t^S}\,dt\Big].
\end{aligned}
\tag{5.43}
$$

We estimate the three terms on the right-hand side of (5.43) separately.

For the first term, we use that $\mathbb{E}_{\hat{\mathbb{Q}}}[\int_0^{\tau^{\Delta\varphi}} c_t^S(\Delta\varphi_t^\lambda)^2\,dt] = O(\lambda^{\frac{2}{p+2}})$ by Proposition 4.4, that $\tau^{\lambda,\mathrm{dual}}$ converges in probability to $\tau^{\Delta\varphi}$ as $\lambda \to 0$ by Lemma D.2, and that $|\Delta\hat{Z}^\lambda| \le \lambda^{\kappa_6}$ on $[\![0, \tau^{\lambda,\mathrm{dual}}]\!]$ by (5.11). Together with dominated convergence, it follows that

$$
\begin{aligned}
&\mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_0^{\tau^{\lambda,\mathrm{dual}}} (1+\Delta\hat{Z}_t^\lambda)\gamma(\Delta\varphi_t^\lambda)^2 c_t^S\,dt\Big]\\
&\qquad = \gamma\mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_0^{\tau^{\Delta\varphi}} c_t^S(\Delta\varphi_t^\lambda)^2\,dt\Big] + o\big(\lambda^{\frac{2}{p+2}}\big).
\end{aligned}
\tag{5.44}
$$

For the second term, we use the definition of ζ^λ in (5.22), the estimate $\lambda^{-\eta}|\Delta S^\lambda_{\tau^{\lambda,\mathrm{dual}}}| \le \lambda^{\kappa_7}$ implied by (5.12), Lemma B.1, and the Cauchy–Schwarz inequality. Together with Lemma 5.3, this yields

$$
\begin{aligned}
&\Big|\mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_{\tau^{\lambda,\mathrm{dual}}}^{\tau^{\lambda,\mathrm{dual}}+\lambda^\eta} \hat{Z}_t^\lambda\zeta_t^\lambda\Delta\varphi_t^\lambda\sqrt{c_t^S}\,dt\Big]\Big|\\
&\qquad \le \mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_{\tau^{\lambda,\mathrm{dual}}}^{\tau^{\lambda,\mathrm{dual}}+\lambda^\eta} \hat{Z}_t^\lambda\lambda^{-\eta}|\Delta S^\lambda_{\tau^{\lambda,\mathrm{dual}}}||\Delta\varphi_t^\lambda|\,dt\Big]\\
&\qquad \le 2T\lambda^{\kappa_7+\eta}\mathbb{E}_{\hat{\mathbb{Q}}}[(\hat{\varphi}_T^*)^2]^{\frac{1}{2}}\mathbb{E}_{\hat{\mathbb{Q}}}[((\hat{Z}^\lambda)^2)_T^*]^{\frac{1}{2}} = o\big(\lambda^{\frac{2}{p+2}}\big).
\end{aligned}
\tag{5.45}
$$

For the third term, we use that $|\Delta\varphi^\lambda||\zeta^\lambda + \gamma\Delta\varphi^\lambda\sqrt{c^S}| \le \lambda^{\kappa_1+\kappa_{10}}$ on $[\![0, \tau^{\lambda,\mathrm{dual}}[\![$ by (3.10) and (5.15) and that $|\Delta\hat{Z}^\lambda| \le \lambda^{\kappa_6}$ on $[\![0, \tau^{\lambda,\mathrm{dual}}]\!]$ by (5.11). This gives

$$\begin{aligned}
&\left|\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}} (1+\Delta\hat{Z}_t^\lambda)\Delta\varphi_t^\lambda(\zeta_t^\lambda + \gamma\Delta\varphi_t^\lambda\sqrt{c^S}_t)\sqrt{c_t^S}\,dt\right]\right| \\
&\quad \le T(1+\lambda^{\kappa_6})\lambda^{\kappa_1+\kappa_{10}}\mathbb{E}_{\hat{\mathbb{Q}}}\big[(\sqrt{c^S})_T^*\big] \\
&\quad = o\big(\lambda^{\frac{2}{p+2}}\big).
\end{aligned} \tag{5.46}$$

Putting (5.44)–(5.45) together now yields the asserted estimate. $\square$

LEMMA 5.8. *Suppose Assumptions 1, 2 and 3 are satisfied. Then*

$$\mathbb{E}_{\hat{\mathbb{Q}}}\big[(\Delta\hat{Z}_T^\lambda)^2\big] = \gamma^2\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}} c_t^S(\Delta\varphi_t^\lambda)^2\,dt\right] + o\big(\lambda^{\frac{2}{p+2}}\big).$$

PROOF. Recall that $\Delta\hat{Z}_T^\lambda$ is a square-integrable $\hat{\mathbb{Q}}$-martingale with quadratic variation

$$\langle\Delta\hat{Z}^\lambda\rangle_T = \int_0^{\tau^{\lambda,\mathrm{dual}}+\lambda^\eta} (1+\Delta\hat{Z}_t^\lambda)^2(\zeta_t^\lambda)^2\,dt.$$

Thus,

$$\begin{aligned}
\mathbb{E}_{\hat{\mathbb{Q}}}\big[(\Delta\hat{Z}_T^\lambda)^2\big] &= \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}} (1+\Delta\hat{Z}_t^\lambda)^2(\zeta_t^\lambda)^2\,dt\right] \\
&\quad + \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_{\tau^{\lambda,\mathrm{dual}}}^{\tau^{\lambda,\mathrm{dual}}+\lambda^\eta} (1+\Delta\hat{Z}_t^\lambda)^2(\zeta_t^\lambda)^2\,dt\right] \\
&= \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}} (1+\Delta\hat{Z}_t^\lambda)^2\gamma^2(\Delta\varphi_t^\lambda)^2 c_t^S\,dt\right] \\
&\quad + \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}} (1+\Delta\hat{Z}_t^\lambda)^2\big((\zeta_t^\lambda)^2 - \gamma^2(\Delta\varphi_t^\lambda)^2 c_t^S\big)\,dt\right] \\
&\quad + \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_{\tau^{\lambda,\mathrm{dual}}}^{\tau^{\lambda,\mathrm{dual}}+\lambda^\eta} (\hat{Z}_t^\lambda)^2(\zeta_t^\lambda)^2\,dt\right].
\end{aligned} \tag{5.47}$$

We estimate the three terms on the right-hand side of (5.47) separately.

For the first term, we use that $\mathbb{E}_{\hat{\mathbb{Q}}}[\int_0^{\tau^{\Delta\varphi}} c_t^S(\Delta\varphi_t^\lambda)^2\,dt] = O(\lambda^{\frac{2}{p+2}})$ by Proposition 4.4, that $\tau^{\lambda,\mathrm{dual}}$ converges in probability to $\tau^{\Delta\varphi}$ as $\lambda \to 0$ by Lemma D.2, and that $|\Delta\hat{Z}^\lambda| \le \lambda^{\kappa_6}$ on $[\![0, \tau^{\lambda,\mathrm{dual}}]\!]$ by (5.11). Together with dominated convergence, it follows that

$$\begin{aligned}
&\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}} (1+\Delta\hat{Z}_t^\lambda)^2\gamma^2(\Delta\varphi_t^\lambda)^2 c_t^S\,dt\right] \\
&\quad = \gamma^2\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}} c_t^S(\Delta\varphi_t^\lambda)^2\,dt\right] + o\big(\lambda^{\frac{2}{p+2}}\big).
\end{aligned} \tag{5.48}$$

For the second term, we use that by (5.15) and (3.10),

$$\begin{aligned}
\big|(\zeta^\lambda)^2 - \gamma^2(\Delta\varphi^\lambda)^2 c^S\big| &\le |\zeta^\lambda + \gamma\Delta\varphi^\lambda\sqrt{c^S}|^2 + 2|\zeta^\lambda + \gamma\Delta\varphi^\lambda\sqrt{c^S}||\gamma\Delta\varphi^\lambda\sqrt{c^S}| \\
&\le \lambda^{2\kappa_{10}} + 2\gamma\sqrt{c^S}\lambda^{\kappa_{10}+\kappa_1} \quad \text{on } [\![0, \tau^{\lambda,\mathrm{dual}}[\![.
\end{aligned}$$

Together with $|\Delta\hat{Z}^\lambda| \leq \lambda^{\kappa_6}$ on $[\![0, \tau^{\lambda,\mathrm{dual}}]\!]$ by (5.11) and the observation that $2\kappa_1 \wedge (\kappa_{10} + \kappa_1) > \frac{2}{2+p}$, this yields

$$\begin{aligned}&\left|\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}} (1+\Delta\hat{Z}_t^\lambda)^2\big((\zeta_t^\lambda)^2 - \gamma^2(\Delta\varphi_t^\lambda)^2 c_t^S\big)\,dt\right]\right| \\ &\quad\leq (1+\lambda^{\kappa_6})^2 T\big(\lambda^{2\kappa_{10}} + 2\gamma\lambda^{\kappa_{10}+\kappa_1}\mathbb{E}_{\hat{\mathbb{Q}}}[(\sqrt{c^S})_T^*]\big) = o\big(\lambda^{\frac{2}{p+2}}\big).\end{aligned} \tag{5.49}$$

For the third term, we use the definition of ζ^λ in (5.22), the estimate that $\lambda^{-\eta}|\Delta S^\lambda_{\tau^{\lambda,\mathrm{dual}}}| \leq \lambda^{\kappa_7}$ by (5.12), Lemma B.1, and the Cauchy–Schwarz inequality. Together with Lemma 5.3, this yields

$$\begin{aligned}&\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_{\tau^{\lambda,\mathrm{dual}}}^{\tau^{\lambda,\mathrm{dual}}+\lambda^\eta} (\hat{Z}_t^\lambda)^2(\zeta_t^\lambda)^2\,dt\right] \\ &\quad\leq \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_{\tau^{\lambda,\mathrm{dual}}}^{\tau^{\lambda,\mathrm{dual}}+\lambda^\eta} (\hat{Z}_t^\lambda)^2\lambda^{-2\eta}|\Delta S^\lambda_{\tau^{\lambda,\mathrm{dual}}}|^2\frac{1}{c_t^S}\,dt\right] \\ &\quad\leq \lambda^{2\kappa_7+\eta}\mathbb{E}_{\hat{\mathbb{Q}}}\big[\big((\hat{Z}^\lambda)^4\big)_T^*\big]^{\frac12}\mathbb{E}_{\hat{\mathbb{Q}}}\big[\big((c^S)^{-2}\big)_T^*\big]^{\frac12} = o\big(\lambda^{\frac{2}{p+2}}\big).\end{aligned} \tag{5.50}$$

The claim now follows by putting together (5.48)–(5.50). □

LEMMA 5.9. *Suppose Assumptions 1, 2 and 3 are satisfied. Then*

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\frac{|\Delta\hat{Z}_T^\lambda|^3}{(\hat{Z}_T^\lambda)^2}\right] = o\big(\lambda^{\frac{2}{p+2}}\big).$$

PROOF. To prove this estimate, we proceed in three steps. First, we apply Itô's formula to $f(\Delta\hat{Z}_T^\lambda)$ where $f = |x|^3/(1+x)^2$ and write the argument of the expectation as a sum of two integrals. We then prove that the stochastic integral in this decomposition is a martingale so that its expectation vanishes. Finally, we show that the Riemann integral is of the right order in λ. To carry out this program, first observe that

$$f'(x) = \frac{3x|x|}{(1+x)^2} - \frac{2|x|^3}{(1+x)^3}, \qquad f''(x) = \frac{6|x|}{(1+x)^2} - \frac{12x|x|}{(1+x)^3} + \frac{6|x|^3}{(1+x)^4}.$$

Itô's formula applied to $f(\Delta\hat{Z}_\cdot^\lambda)$ in turn gives

$$\begin{aligned}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\frac{|\Delta\hat{Z}_T^\lambda|^3}{(\hat{Z}_T^\lambda)^2}\right] &= \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T f'(\Delta\hat{Z}_t^\lambda)\,d\Delta\hat{Z}_t^\lambda\right] \\ &\quad+ \frac12\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T f''(\Delta\hat{Z}_t^\lambda)\,d\langle\Delta\hat{Z}^\lambda\rangle_t\right] \\ &= -\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T f'(\Delta\hat{Z}_t^\lambda)\hat{Z}_t^\lambda\zeta_t^\lambda\,dW_t^{S,\hat{\mathbb{Q}}}\right] \\ &\quad+ \frac12\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T f''(\Delta\hat{Z}_t^\lambda)(\hat{Z}_t^\lambda)^2(\zeta_t^\lambda)^2\,dt\right].\end{aligned} \tag{5.51}$$

To prove that the stochastic integral on the right-hand side is a martingale, we use the elementary inequality $(a+b)^2 \leq 2(a^2+b^2)$ to obtain

$$\begin{aligned}&\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T f'(\Delta\hat{Z}_t^\lambda)^2(\hat{Z}_t^\lambda)^2(\zeta_t^\lambda)^2\,dt\right] \\ &\quad\leq 18\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T \frac{(\Delta\hat{Z}_t^\lambda)^4}{(\hat{Z}_t^\lambda)^2}(\zeta_t^\lambda)^2\,dt\right] + 8\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T \frac{(\Delta\hat{Z}_t^\lambda)^6}{(\hat{Z}_t^\lambda)^4}(\zeta_t^\lambda)^2\,dt\right].\end{aligned}$$

By Lemma 5.10, the expectations on the right-hand side are indeed finite.

It therefore remains then to show that the second expectation of the Riemmann integral in (5.51) is of the right order in λ. By the triangle inequality, it suffices to prove that $A_5 + A_6 + A_7 = o(\lambda^{\frac{2}{p+2}})$, where

$$A_5 := \mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_0^T |\Delta \hat{Z}_t^\lambda| (\zeta_t^\lambda)^2 \, dt\Big], \qquad A_6 := \mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_0^T \frac{(\Delta \hat{Z}_t^\lambda)^2}{\hat{Z}_t^\lambda} (\zeta_t^\lambda)^2 \, dt\Big],$$

$$A_7 := \mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_0^T \frac{|\Delta \hat{Z}_t^\lambda|^3}{(\hat{Z}_t^\lambda)^2} (\zeta_t^\lambda)^2 \, dt\Big].$$

This holds by Lemma 5.10 below, so that the proof is complete. □

LEMMA 5.10. *Let $k \geq 1$ and $\ell \geq 0$. Then*

$$\mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_0^T \frac{|\Delta \hat{Z}_t^\lambda|^k}{(\hat{Z}_t^\lambda)^\ell} (\zeta_t^\lambda)^2 \, dt\Big] = o(\lambda^{\frac{2}{p+2}}).$$

PROOF. Using the definition of ζ^λ in (5.22), we obtain

$$\begin{aligned} &\mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_0^T \frac{|\Delta \hat{Z}_t^\lambda|^k}{(\hat{Z}_t^\lambda)^\ell} (\zeta_t^\lambda)^2 \, dt\Big] \\ &\quad = \mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_0^{\tau^{\lambda,\mathrm{dual}}} \frac{|\Delta \hat{Z}_t^\lambda|^k}{(\hat{Z}_t^\lambda)^\ell} (\zeta_t^\lambda)^2 \, dt\Big] \\ &\qquad + \mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_{\tau^{\lambda,\mathrm{dual}}}^{\tau^{\lambda,\mathrm{dual}}+\lambda^\eta} \frac{|\Delta \hat{Z}_t^\lambda|^k}{(\hat{Z}_t^\lambda)^\ell} \frac{(\lambda^{-\eta} \Delta S^\lambda_{\tau^{\lambda,\mathrm{dual}}})^2}{c_t^S} \, dt\Big]. \end{aligned} \tag{5.52}$$

We estimate the two terms on the right-hand side of (5.52) separately.

For the first term, we use that $|\zeta^\lambda| \vee |\Delta \hat{Z}^\lambda| \leq \lambda^{\kappa_6}$ on $[\![0, \tau^{\lambda,\mathrm{dual}}[\![$ by (5.10) and (5.11). Together with $k \geq 1$, Lemma 5.3 and the fact that $3\kappa_6 > \frac{2}{p+2}$, this yields

$$\mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_0^{\tau^{\lambda,\mathrm{dual}}} \frac{|\Delta \hat{Z}_t^\lambda|^k}{(\hat{Z}_t^\lambda)^\ell} (\zeta_t^\lambda)^2 \, dt\Big] \leq \lambda^{3\kappa_6} T \mathbb{E}_{\hat{\mathbb{Q}}}[((\hat{Z}^\lambda)^{-\ell})_T^*] = o(\lambda^{\frac{2}{p+2}}). \tag{5.53}$$

For the second term, we use the definition of ζ^λ in (5.22), the estimate that $\lambda^{-\eta}|\Delta S^\lambda_{\tau^{\lambda,\mathrm{dual}}}| \leq \lambda^{\kappa_7}$ by (5.12), the elementary inequality $|a-1|^k \leq 2^k(a^k+1)$ for $a \geq 0$, and the Cauchy–Schwarz inequality. Together with Lemma 5.3, this shows

$$\begin{aligned} &\mathbb{E}_{\hat{\mathbb{Q}}}\Big[\int_{\tau^{\lambda,\mathrm{dual}}}^{\tau^{\lambda,\mathrm{dual}}+\lambda^\eta} \frac{|\Delta \hat{Z}_t^\lambda|^k}{(\hat{Z}_t^\lambda)^\ell} \frac{(\lambda^{-\eta} \Delta S^\lambda_{\tau^{\lambda,\mathrm{dual}}})^2}{c_t^S} \, dt\Big] \\ &\quad \leq 2^k T \lambda^{2\kappa_7+\eta} \mathbb{E}_{\hat{\mathbb{Q}}}[(((\hat{Z}^\lambda)^{k-\ell} + (\hat{Z}^\lambda)^{-\ell})(c^S)^{-1})_T^*] \\ &\quad \leq 2^k T \lambda^{2\kappa_7+\eta} \mathbb{E}_{\hat{\mathbb{Q}}}[((c^S)^{-2})_T^*]^{\frac{1}{2}} \\ &\qquad \times (\mathbb{E}_{\hat{\mathbb{Q}}}[((\hat{Z}^\lambda)^{2(k-\ell)})_T^*]^{\frac{1}{2}} + \mathbb{E}_{\hat{\mathbb{Q}}}[((\hat{Z}^\lambda)^{-2\ell})_T^*]^{\frac{1}{2}}) \\ &\quad = o(\lambda^{\frac{2}{p+2}}). \end{aligned} \tag{5.54}$$

The claim now follows by putting together (5.53) and (5.54). □

APPENDIX A: RESULTS ON THE FUNCTION g_p

In this Appendix, we prove some auxiliary results on the function g_p from (3.1) that are used in the proof of Theorem 3.3.

COROLLARY A.1.

(i) *For every* $r \in (0, \infty)$, *there exists a constant* $C > 0$ *such that*

$$|\tilde{g}_p(z)|^r = |g_p(z)|^{\frac{r}{p-1}} \leq C(|z|^{2r} + 1), \quad z \in \mathbb{R}. \tag{A.1}$$

(ii) *There exists a constant* $C > 0$ *such that*

$$|z\tilde{g}_p(z)| \geq C(|z|^{\frac{2+p}{p}} - 1), \quad z \in \mathbb{R}, \tag{A.2}$$

$$|z\tilde{g}_p(z)| \geq C|z|^2 \mathbb{1}_{\{|z|\geq 1\}}, \quad z \in \mathbb{R}. \tag{A.3}$$

(iii) *There exists a constant* $C > 0$ *such that*

$$|g_p'(z)| \leq C(|z|^2 + 1), \quad z \in \mathbb{R}.$$

PROOF. (i) This follows from the continuity of g_p, its growth rate at infinity (3.2), and $2/p \in (1, 2)$.

(ii) Recall from Lemma 3.1 that $g_p(0) = 0$ and $c_p > 0$. As g_p satisfies (3.1), it follows that $g_p(z) > \frac{c_p}{2} z$ on a sufficiently small interval $[0, \delta]$ with $\delta > 0$. On $(0, \sqrt{c_p}]$, this gives $g_p'(z) = -z^2 + c_p + (p-1)p^{-\frac{p}{p-1}}|g_p(z)|^{\frac{p}{p-1}} > 0$. Thus, on $[\delta, \sqrt{c_p}]$ the function $z \mapsto g_p(z)/z$ is continuous and bounded from below by a constant C. As g_p is odd, we have $|g_p(z)| \geq (C \wedge c_p/2)|z|$ on $[-\sqrt{c_p}, \sqrt{c_p}]$.

The growth condition for g_p at infinity gives the existence of constants C' and $K \geq 1$ such that $|g_p(z)| \geq C'|z|^{\frac{2(p-1)}{p}}$ for $|z| \geq K$. Moreover, on $[\sqrt{c_p}, K]$ it holds that $g_p'(z) \geq 0$. Therefore, $|g_p(z)|/|z|$ is bounded from below on $(0, \sqrt{c_p}]$ by $C \wedge c_p/2$, is continuous and bounded from below by $g_p(\sqrt{c_p})/K$ on $[\sqrt{c_p}, K]$. This implies that the following holds for some $C'' \leq (C \wedge c_p/2 \wedge C')^{\frac{1}{p-1}}$:

$$|\tilde{g}_p(z)| = |g_p(z)|^{\frac{1}{p-1}} \geq C''|z|^{\frac{1}{p-1}}, \quad z \in [-K, K], \tag{A.4}$$

$$|\tilde{g}_p(z)| = |g_p(z)|^{\frac{1}{p-1}} \geq C''|z|^{\frac{2}{p}}, \quad z \in \mathbb{R}\backslash[-K, K]. \tag{A.5}$$

As a result, the function $|g_p(z)|/|z|^{\frac{2(p-1)}{p}}$ is strictly positive, continuous and, therefore, bounded from below on $\mathbb{R}\backslash[-1, 1]$. Similarly, $|g_p(z)|/|z|$ is strictly positive, continuous and, therefore, bounded from below on $[-1, 1]\backslash\{0\}$. This gives the inequalities (A.2) and (A.3).

To show that $g_p'(z) \geq 0$ on $[\sqrt{c_p}, K]$, assume to the contrary that there is a z_1 in the interval such that $g_p'(z_1) < 0$. Then for all $z \geq z_1$ with $g_p(z) \geq -p|\frac{1}{p-1}(z^2 - c_p)|^{\frac{p-1}{p}}$, it holds that $g_p'(z) < 0$, which contradicts the growth condition for g_p at $+\infty$.

(iii) This follows from the ODE (3.1), the triangle inequality and the growth conditions (3.2). □

APPENDIX B: BOUNDING THE CANDIDATE STRATEGY BY THE FRICTIONLESS OPTIMIZER

The following estimate allows to deduce the existence of moments for the displacement $\Delta\varphi_t^\lambda$ from the corresponding integrability of the supremum of the frictionless optimal strategy. This is used in the proof that our candidate strategy is admissible (see Theorem 3.3) and in the estimation of the primal and dual bounds.

LEMMA B.1. *Suppose that* $|\varphi_0^\lambda| \leq |\hat{\varphi}_0|$. *Then the candidate strategy* φ_t^λ *satisfies*

$$|\varphi_t^\lambda| \leq \hat{\varphi}_t^* \quad \forall t \in [0, T].$$

As a consequence,

$$|\hat{\varphi}_t - \varphi_t^\lambda| \leq 2\hat{\varphi}_t^* \quad \forall t \in [0, T]. \tag{B.1}$$

PROOF. It suffices to show that $|\varphi^\lambda| \leq \hat{\varphi}^*$ on $[\![0, \tau^{\Delta\varphi}[\![$ since $|\varphi^\lambda| \leq |\varphi^\lambda_{\tau^{\Delta\varphi}}|$ on $[\![\tau^{\Delta\varphi}, T]\!]$ by definition of φ^λ.

Fix $\omega \in \Omega$ and let $\tau_0 = \inf\{t \in [0, \tau^{\Delta\varphi}] | |\varphi_t^\lambda| > \hat{\varphi}_t^*\} \wedge \tau^{\Delta\varphi}$. Note that $\tau_0 = \tau^{\Delta\varphi}$ or $\tau_0 < \tau^{\Delta\varphi}$. We want to show that $\tau_0 = \tau^{\Delta\varphi}$ (and consequently that $|\varphi_t^\lambda| \leq \hat{\varphi}_t^*, \forall t \in [0, T]$). We assume that $\tau_0 \in [0, \tau^{\Delta\varphi})$ and want to obtain a statement contradicting the assumptions made.

By continuity of φ^λ and $\hat{\varphi}^*$, we have $|\varphi_{\tau_0}^\lambda| = \hat{\varphi}_{\tau_0}^*$. Furthermore, by definition of the infimum, there exist $\varepsilon > 0$ and $\tau_1 \in (\tau_0, \tau^{\Delta\varphi})$ such that $|\varphi_{\tau_1}^\lambda| > \hat{\varphi}_{\tau_1}^* + \varepsilon$. Let $\tau_2 = \inf\{t \in [0, \tau_1] | |\varphi_t^\lambda| > \hat{\varphi}_t^* + \frac{\varepsilon}{2}\}$. By continuity of $\hat{\varphi}^*$ and φ^λ and the definition of τ_0 and τ_1, it holds that $\tau_0 < \tau_2 < \tau_1$. We now prove that the definition of the trading rate $\dot{\varphi}^\lambda$ and the definition of a derivative for φ^λ contradict the assumption that $\tau_0 \in [0, \tau^{\Delta\varphi})$.

Without loss of generality, we can assume that $\varphi_{\tau_2}^\lambda > \hat{\varphi}_{\tau_2}^* \geq \hat{\varphi}_{\tau_2}$ (the case where $-\varphi_{\tau_2}^\lambda > \hat{\varphi}_{\tau_2}^* \geq \hat{\varphi}_{\tau_2}$ is treated similarly), which implies by definition of $\dot{\varphi}^\lambda$ (see (3.6)) that $\dot{\varphi}_{\tau_2}^\lambda = \lim_{t\to\tau_2^+} \frac{\varphi_t^\lambda - \varphi_{\tau_2}^\lambda}{t-\tau_2} < 0$. However, by definition of τ_2, for every $\delta > 0$ there exists $\tau^\delta \in (\tau_2, \tau_2 + \delta)$ such that $\varphi_{\tau^\delta}^\lambda > \varphi_{\tau_2}^\lambda$. This contradicts the existence of a negative limit. $\square$

APPENDIX C: PRIMAL STOPPING-TIME BOUNDS

In this Appendix, we prove Proposition C.5, which is a key ingredient for the proof of Proposition 4.1. It estimates the effect of stopping trading too early, that is, at $\tau^{\Delta\varphi}$ instead of T^λ.

For better readability, the proof of Proposition C.5 is broken up into four lemmas. The crucial one is Lemma C.2 which establishes a maximal inequality in the spirit of Peskir [52] for the process $\Delta\varphi^\lambda$. This is a delicate matter, since the mean-reversion speed of this process is not bounded from below. However, this can be overcome using a result established in the companion paper of the present study [15].

Note that the stopping time $\tau^{\Delta\varphi}$ defined in (3.10) can be rewritten as the minimum of the following five stopping times:

$$\tau^{\Delta\varphi} = \tau^{\lambda,m,1} \wedge \tau^{\lambda,m,2} \wedge \tau^{\lambda,\Delta\varphi} \wedge \tau^{\lambda,\mathrm{cost}} \wedge \tau^{\lambda,\hat{\varphi}},$$

where

$$\tau^{\lambda,m,1} = \inf\left\{t \in [0, T^\lambda] : \frac{1}{2} p^{-\frac{1}{p-1}} m_t < \lambda^{\kappa_2}\right\} \wedge T^\lambda, \tag{C.1}$$

$$\tau^{\lambda,m,2} = \inf\{t \in [0, T^\lambda] : m_t > \lambda^{-\kappa_2}\} \wedge T^\lambda, \tag{C.2}$$

$$\tau^{\lambda,\hat{\varphi}} = \inf\{t \in [0, T^\lambda] : |\hat{\varphi}_t| \geq \lambda^{-\kappa_4}\} \wedge T^\lambda, \tag{C.3}$$

$$\tau^{\lambda,\Delta\varphi} = \inf\{t \in [0, \tau^{\lambda,m,1} \wedge \tau^{\lambda,m,2}] | |\overline{\Delta\varphi}^\lambda| > \lambda^{\kappa_1}\} \wedge \tau^{\lambda,m,1} \wedge \tau^{\lambda,m,2}, \tag{C.4}$$

$$\tau^{\lambda,\mathrm{cost}} = \inf\{t \in [0, \tau^{\lambda,\Delta\varphi}] : \lambda\Lambda_t |\dot{\overline{\varphi}}_t^\lambda|^p > \lambda^{\kappa_3}\} \wedge \tau^{\lambda,\Delta\varphi}. \tag{C.5}$$

We proceed to show that all these five stopping times converge in $\hat{\mathbb{Q}}$-probability to T^λ sufficiently fast (in the order $O(\lambda^{\frac{4(1+2\varepsilon)}{p+2}})$). As stated above, Lemma C.2 is the key mathematical ingredient.

LEMMA C.1. *Suppose Assumptions 1, 2 and 3 are satisfied. Then*

$$\hat{\mathbb{Q}}[\tau^{\lambda,m,1} < T^\lambda] + \hat{\mathbb{Q}}[\tau^{\lambda,m,2} < T^\lambda] + \hat{\mathbb{Q}}[\tau^{\lambda,\hat{\varphi}} < T^\lambda] = O(\lambda^{\frac{4(1+2\varepsilon)}{p+2}}).$$

PROOF. The assertion follows immediately from Markov's inequality. □

LEMMA C.2. *Suppose Assumptions 1, 2 and 3 are satisfied. Then*

$$\mathbb{E}_{\hat{\mathbb{Q}}}\Big[\max_{0\leq t\leq \tau^{\lambda,m,1}\wedge\tau^{\lambda,m,2}} (\lambda^{-\frac{1}{p+2}} \overline{\Delta\varphi}_t^\lambda)^n\Big] = o(\lambda^{-6n\kappa_2}) \quad \textit{for } n \in \mathbb{N} \textit{ with } n \geq 1.$$

PROOF. We apply [15], Lemma B.4, with $\varepsilon = \lambda^{\frac{1}{p+2}}$, $b = \mu^{\hat{\varphi}}$, $c = c^{\hat{\varphi}}$, $M = m$, $L = \frac{1}{2} p^{-\frac{1}{p-1}} m$, $\kappa = \lambda^{\kappa_2}$. Note that the functions C^1 and C^2 from [15], Proposition B.1, satisfy

$$C^1(\gamma, \sigma, n) = O(\gamma^{-4n-1}) \leq O(\gamma^{-5n}) \quad \text{and}$$
$$C^2(\gamma, n) = O(\gamma^{-3n-1}) \leq O(\gamma^{-4n})$$

for fixed $\sigma > 0$ and $n \in \mathbb{N}$. Thus, for some constants $a, C > 0$,[26] we obtain

$$\begin{aligned}
&\mathbb{E}_{\hat{\mathbb{Q}}}\Big[\max_{0\leq t\leq \tau^{\lambda,m,1}\wedge\tau^{\lambda,m,2}} (\lambda^{-\frac{1}{p+2}} \overline{\Delta\varphi}_t^\lambda)^n\Big] \\
&\quad \leq C\bigg(\sqrt{C^1(a\lambda^{2\kappa_2}, 2, n)} + \sqrt{C^2(a\lambda^{2\kappa_2}, n)} \\
&\qquad \times \mathbb{E}_{\hat{\mathbb{Q}}}\bigg[\log\bigg(1 \vee \lambda^{-\frac{2}{p+2}} \int_0^{\tau^{\lambda,m,1}\wedge\tau^{\lambda,m,2}} c_t^{\hat{\varphi}}\, dt\bigg)^n\bigg]^{\frac{1}{2}}\bigg) \\
&\quad = O\bigg(\lambda^{-(4n+1)\kappa_2} \log\bigg(\frac{1}{\lambda}\bigg)\bigg) = o(\lambda^{-6n\kappa_2}),
\end{aligned}$$

as asserted. □

LEMMA C.3. *Suppose Assumptions 1, 2 and 3 are satisfied. Then*

$$\hat{\mathbb{Q}}[\tau^{\lambda,\Delta\varphi} < \tau^{\lambda,m,1} \wedge \tau^{\lambda,m,2}] = O(\lambda^{\frac{4(1+2\varepsilon)}{p+2}}). \tag{C.6}$$

PROOF. Observe that $\frac{1}{p+2} - \kappa_1 - 6\kappa_2 > 0$. Markov's inequality and Lemma C.2 for $n \in \mathbb{N}$ with $n \geq \frac{4(1+2\varepsilon)}{(p+2)(\frac{1}{p+2}-\kappa_1-6\kappa_2)}$ in turn yield the desired estimate:

$$\hat{\mathbb{Q}}[\tau^{\lambda,\Delta\varphi} < \tau^{\lambda,m,1} \wedge \tau^{\lambda,m,2}] \leq \hat{\mathbb{Q}}\Big[\sup_{0\leq t\leq \tau^{\lambda,m,1}\wedge\tau^{\lambda,m,2}} |\overline{\Delta\varphi}_t^\lambda| > \lambda^{\kappa_1}\Big]$$

[26] Note that by the first moment of Assumption 3 and Novikov's criterion $\mathbb{E}[d\mathbb{P}/d\hat{\mathbb{Q}}] < \infty$.

$$\leq \lambda^{n(\frac{1}{p+2}-\kappa_1)}\mathbb{E}_{\hat{\mathbb{Q}}}\Big[\max_{0\leq t\leq \tau^{\lambda,m,1}\wedge\tau^{\lambda,m,2}}\big(\lambda^{-\frac{1}{p+2}}\overline{\Delta\varphi}_t^{\lambda}\big)^n\Big]$$

$$= O\big(\lambda^{n(\frac{1}{p+2}-\kappa_1-6\kappa_2)}\big) \leq O\big(\lambda^{\frac{4(1+2\varepsilon)}{p+2}}\big). \qquad \square$$

LEMMA C.4. *Suppose Assumptions 1, 2 and 3 are satisfied. Then*

$$\hat{\mathbb{Q}}\big[\tau^{\lambda,\mathrm{cost}} < \tau^{\lambda,\Delta\varphi}\big] = O\big(\lambda^{\frac{4(1+2\varepsilon)}{p+2}}\big).$$

PROOF. We use the ODE (3.6) for φ^λ, Corollary A.1, $|\overline{\Delta\varphi}_t^{\lambda}| \leq \lambda^{\kappa_1}$ on $[0, \tau^{\lambda,\Delta\varphi}]$, and $\kappa_1 < 1/(p+2)$. For $\lambda \leq 1$ and some constant $C > 0$, this yields

$$\begin{aligned}|\dot{\varphi}_t^{\lambda}| &\leq C\lambda^{-\frac{1}{p+2}}c_t^{\hat{\varphi}}m_t\big(1 + m_t^2\lambda^{-\frac{2}{p+2}}\big(\overline{\Delta\varphi}_t^{\lambda}\big)^2\big) \leq C\lambda^{-\frac{1}{p+2}}c_t^{\hat{\varphi}}m_t\big(1 + m_t^2\lambda^{-\frac{2}{p+2}+2\kappa_1}\big)\\ &\leq C\lambda^{-\frac{3}{p+2}+2\kappa_1}c_t^{\hat{\varphi}}\big(1 + m_t^3\big) \leq C\lambda^{-\frac{3}{p+2}+2\kappa_1}\big(c_T^{\hat{\varphi}*}\big(1 + \big(m_T^*\big)^3\big)\big) \quad \text{on } [\![0, \tau^{\lambda,\Delta\varphi}]\!].\end{aligned}$$

Since $2 - 2p + (p+2)(2p\kappa_1 - \kappa_3) > 0$, the inequalities of Markov and Hölder with Assumption 3 now yield the asserted estimate:

$$\begin{aligned}&\hat{\mathbb{Q}}\big[\tau^{\lambda,\mathrm{cost}} < \tau^{\lambda,\Delta\varphi}\big]\\ &\quad\leq \hat{\mathbb{Q}}\big[\big(\lambda\Lambda|\dot{\varphi}^{\lambda}|^p\big)^*_{\tau^{\lambda,\Delta\varphi}} \geq \lambda^{\kappa_3}\big]\\ &\quad\leq \hat{\mathbb{Q}}\big[C\lambda^{1-\frac{3p}{p+2}+2p\kappa_1}\Lambda_T^*\big(c_T^{\hat{\varphi}*}\big(1+\big(m_T^*\big)^3\big)\big)^p \geq \lambda^{\kappa_3}\big]\\ &\quad\leq C\lambda^{\frac{4(1+2\varepsilon)}{p+2}}\mathbb{E}_{\hat{\mathbb{Q}}}\Big[\big(\Lambda_T^*\big)^{\frac{4(1+2\varepsilon)}{2-2p+(p+2)(2p\kappa_1-\kappa_3)}}\big(c_T^{\hat{\varphi}*}\big(1+\big(m_T^*\big)^3\big)\big)^{\frac{4(1+2\varepsilon)p}{2-2p+(p+2)(2p\kappa_1-\kappa_3)}}\Big]\\ &\quad= C\lambda^{\frac{4(1+\varepsilon)}{p+2}}\mathbb{E}_{\hat{\mathbb{Q}}}\Big[\big(\Lambda_T^*\big)^{\frac{4(1+2\varepsilon)(1+\frac{1}{\varepsilon})}{2-2p+(p+2)(2p\kappa_1-\kappa_3)}}\Big]^{\frac{1}{1+\frac{1}{\varepsilon}}}\\ &\qquad\times\mathbb{E}_{\hat{\mathbb{Q}}}\Big[\big(c_T^{\hat{\varphi}*}\big(1+\big(m_T^*\big)^3\big)\big)^{\frac{4(1+2\varepsilon)(1+\varepsilon)p}{2-2p+(p+2)(2p\kappa_1-\kappa_3)}}\Big]^{\frac{1}{1+\varepsilon}}\\ &\quad= O\big(\lambda^{\frac{4(1+2\varepsilon)}{p+2}}\big).\end{aligned} \qquad \square$$

By combining Lemmas C.1, C.3 and C.4, we finally obtain the following estimate, which is a key ingredient for the proof of Proposition 4.1.

PROPOSITION C.5. *Suppose Assumptions 1, 2 and 3 are satisfied. Then*

$$\hat{\mathbb{Q}}\big[\tau^{\Delta\varphi} < T^{\lambda}\big] = O\big(\lambda^{\frac{4(1+2\varepsilon)}{p+2}}\big). \tag{C.7}$$

PROOF. In view of Lemmas C.1, C.3 and C.4, we have

$$\begin{aligned}\hat{\mathbb{Q}}\big[\tau^{\Delta\varphi} < T^{\lambda}\big] &\leq \hat{\mathbb{Q}}\big[\tau^{\lambda,m,1} < T^{\lambda}\big] + \hat{\mathbb{Q}}\big[\tau^{\lambda,m,2} < T^{\lambda}\big] + \hat{\mathbb{Q}}\big[\tau^{\lambda,\hat{\varphi}} < T^{\lambda}\big]\\ &\quad + \hat{\mathbb{Q}}\big[\tau^{\lambda,\Delta\varphi} < \tau^{\lambda,m}\big] + \hat{\mathbb{Q}}\big[\tau^{\lambda,\mathrm{cost}} < \tau^{\lambda,\Delta\varphi}\big] = O\big(\lambda^{\frac{4(1+2\varepsilon)}{p+2}}\big).\end{aligned} \qquad \square$$

APPENDIX D: DUAL STOPPING TIME BOUNDS

In this section, we prove the convergence in $\hat{\mathbb{Q}}$-probability of $\tau^{\lambda,\mathrm{dual}}$ to $\tau^{\Delta\varphi}$ as $\lambda \to 0$, where the stopping times are defined respectively in (5.16) and (3.10). This result is used in the proofs of Lemmas 5.6–5.8 to bound various dual remainder terms.

LEMMA D.1. *Suppose Assumptions 1, 2 and 3 are satisfied. Then the following limits hold in probability as $\lambda \to 0$:*

(i) $\max_{0\le t\le \tau^{\Delta\varphi}} \lambda^{\kappa_5} B_t^\lambda \longrightarrow 0$,

(ii) $\max_{0\le t\le \tau^{\lambda,1}} \lambda^{-\kappa_6} \left|\frac{\mu_t^{\overline{S}^\lambda,\hat{\mathbb{Q}}}\sqrt{c_t^S}}{c_t^{\overline{S}^\lambda,S}}\right| \longrightarrow 0$,

(iii) $\max_{0\le t\le \tau^{\lambda,2}} \lambda^{-\kappa_6} \left|\mathcal{E}\left(\int_0^\cdot \frac{\mu_t^{\overline{S}^\lambda,\hat{\mathbb{Q}}}\sqrt{c_t^S}}{c_t^{\overline{S}^\lambda,S}}\,dW^{S,\hat{\mathbb{Q}}}\right)_t - 1\right| \longrightarrow 0$,

(iv) $\max_{0\le t\le \tau^{\lambda,1}} \lambda^{-\kappa_7}(\lambda^{-\eta}|\overline{\Delta S}_t^\lambda|) \longrightarrow 0$,

(v) $\max_{0\le t\le \tau^{\lambda,1}} \lambda^{-\kappa_8} \frac{1}{c_t^S}\left|c_t^{\overline{\Delta S}^\lambda} + 2c_t^{S,\overline{\Delta S}^\lambda}\right| \longrightarrow 0$,

(vi) $\max_{0\le t\le T} \lambda^{\kappa_9} \Lambda^{\frac{1}{p-1}} \longrightarrow 0$,

(vii) $\max_{0\le t\le \tau^{\lambda,1}} \lambda^{-\kappa_{10}} \left|\frac{\mu_t^{\overline{S}^\lambda,\hat{\mathbb{Q}}}\sqrt{c_t^S}}{c_t^{\overline{S}^\lambda,S}} + \gamma\sqrt{c_t^S}\Delta\varphi_t^\lambda\right| \longrightarrow 0$.

PROOF. First, observe that Lemma C.2 shows $\mathbb{E}_{\hat{\mathbb{Q}}}[\max_{0\le t\le \tau^{\Delta\varphi}} \lambda^{-\frac{1}{p+2}}\overline{\Delta\varphi}_t^\lambda] = o(\lambda^{-6\kappa_2})$. Together with the fact that $\kappa_5 - 6\kappa_2 > 0$ and the definition of B^λ in (5.4), this implies Item (i). Next, note that Item (i), Corollary A.1, and $p-1\le 1$ give the following limits in probability:

$$\begin{aligned}&\max_{0\le t\le \tau^{\Delta\varphi}} \lambda^{2\kappa_5}\tilde{g}_p(B_t^\lambda) \to 0, \qquad \max_{0\le t\le \tau^{\Delta\varphi}} \lambda^{2\kappa_5} g_p(B_t^\lambda) \to 0,\\ &\max_{0\le t\le \tau^{\Delta\varphi}} \lambda^{2\kappa_5} g_p'(B_t^\lambda) \to 0.\end{aligned} \tag{D.1}$$

Now, the definition and the dynamics (5.7) of $\overline{\Delta S}^\lambda$ and equation (5.8) imply that, on $[\![0,\tau^{\Delta\varphi}]\!]$,

$$|\overline{\Delta S}_t^\lambda| \le \lambda^{\frac{3}{p+2}-2\kappa_5} A_t |\lambda^{2\kappa_5} g_p(B_t^\lambda)|, \tag{D.2}$$

$$|\gamma c_t^S \overline{\Delta\varphi}_t^\lambda| \le \lambda^{\frac{1}{p+2}-\kappa_5} A_t c^{\hat\varphi} m_t^2 |\lambda^{\kappa_5} B_t^\lambda|, \tag{D.3}$$

$$\begin{aligned}|\mu_t^{\overline{S}^\lambda,\hat{\mathbb{Q}}} + \gamma c_t^S \overline{\Delta\varphi}^\lambda| \le{}& \lambda^{\frac{3}{p+2}-6\kappa_2-2\kappa_5} A_t \mu_t^{m,\hat{\mathbb{Q}}} |\lambda^{2\kappa_5} g_p'(B_t^\lambda)||\lambda^{-\frac{1}{p+2}+6\kappa_2}\overline{\Delta\varphi}_t^\lambda|\\ &+ \lambda^{\frac{2}{p+2}-2\kappa_5} A_t m_t \mu_t^{\hat\varphi,\hat{\mathbb{Q}}} |\lambda^{2\kappa_5} g_p'(B_t^\lambda)|\\ &+ \lambda^{\frac{2}{p+2}-2\kappa_5} A_t \sqrt{c_t^{\hat\varphi}}\, c_t^{m,W^{\hat\varphi,\hat{\mathbb{Q}}}} |\lambda^{2\kappa_5} g_p'(B_t^\lambda)|\\ &+ \lambda^{\frac{2}{p+2}-2\kappa_5} m_t \sqrt{c_t^{\hat\varphi}}\, c_t^{A,W^{\hat\varphi,\hat{\mathbb{Q}}}} |\lambda^{2\kappa_5} g_p'(B_t^\lambda)|\\ &+ \lambda^{\frac{3}{p+2}-2\kappa_5} \mu_t^{A,\hat{\mathbb{Q}}} |\lambda^{2\kappa_5} g_p(B_t^\lambda)|,\end{aligned} \tag{D.4}$$

$$\begin{aligned}|c^{\overline{\Delta S}^\lambda,S}| \le{}& \lambda^{\frac{3}{p+2}-6\kappa_2-2\kappa_5} A_t |c_t^{S,m}||\lambda^{2\kappa_5} g_p'(B_t^\lambda)||\lambda^{-\frac{1}{p+2}+6\kappa_2}\overline{\Delta\varphi}_t^\lambda|\\ &+ \lambda^{\frac{2}{p+2}-2\kappa_5} A_t m_t \sqrt{c_t^{\hat\varphi}}\,|c_t^{S,W^{\hat\varphi,\hat{\mathbb{Q}}}}||\lambda^{2\kappa_5} g_p'(B_t^\lambda)|\\ &+ \lambda^{\frac{3}{p+2}-2\kappa_5} |c_t^{S,A}||\lambda^{2\kappa_5} g_p(B_t^\lambda)|,\end{aligned} \tag{D.5}$$

$$\begin{aligned}c_t^{\overline{\Delta S}^\lambda} ={}& \lambda^{\frac{6}{p+2}-4\kappa_5-12\kappa_2} A_t^2 c_t^m (\lambda^{2\kappa_5} g_p'(B_t^\lambda))^2 (\lambda^{-\frac{1}{p+2}+6\kappa_2}\overline{\Delta\varphi}_t^\lambda)^2\\ &+ \lambda^{\frac{4}{p+2}-4\kappa_5} A_t^2 m_t^2 c_t^{\hat\varphi} (\lambda^{2\kappa_5} g_p'(B_t^\lambda))^2\\ &+ \lambda^{\frac{6}{p+2}-4\kappa_5} (\lambda^{2\kappa_5} g_p(B_t^\lambda))^2 c_t^A.\end{aligned} \tag{D.6}$$

Now, (D.2) and (5.18) show that $\max_{0\le t\le \tau^{\lambda,1}} \lambda^{-\kappa_7-\eta}|\overline{\Delta S}^{\lambda}_t| \to 0$ in probability verifying Item (iv). By combining (D.3), (D.4) and (5.17), we obtain $\max_{0\le t\le \tau^{\lambda,1}} \lambda^{-\kappa_6}|\mu^{\overline{S}^{\lambda},\mathbb{Q}}| \to 0$ in probability. Next, observe that (D.4) and (5.19) show $\max_{0\le t\le \tau^{\lambda,1}} \lambda^{-\kappa_{10}}|\mu_t^{\overline{S}^{\lambda},\hat{\mathbb{Q}}} + \gamma c_t^S \overline{\Delta\varphi}^{\lambda}| \to 0$ in probability. Finally, (D.5) and (5.18) give $\max_{0\le t\le \tau^{\lambda,1}} \lambda^{-\kappa_8}|c_t^{\overline{\Delta S}^{\lambda},S}| \to 0$ in probability, and (D.6) together with (5.18) shows $\max_{0\le t\le \tau^{\lambda,1}} \lambda^{-\kappa_8}|c_t^{\overline{\Delta S}^{\lambda}}| \to 0$ in probability. Now, recall that $c^{\overline{S}^{\lambda},S} = c^S + c^{\overline{\Delta S}^{\lambda},S}$. Combining the above in turn establishes Item (ii), (vii) and (v). Item (ii) also implies that

$$\max_{0\le t\le \tau^{\lambda,2}} \lambda^{-\kappa_6}\left|\int_0^t \frac{\mu_t^{\overline{S}^{\lambda},\hat{\mathbb{Q}}}\sqrt{c_t^S}}{c_t^{\overline{S}^{\lambda},S}}\,dW_t^{S,\hat{\mathbb{Q}}}\right| \to 0 \quad \text{in probability.}$$

Together with the elementary inequality $|\exp(x)-1| \le 2x$ for $x\le 1$ and Item (ii), this establishes Item (iii). Finally, the continuity of Λ and $\kappa_9 > 0$ show that Item (vi) holds as well. □

With Lemma D.1 at hand, we can now establish the last missing piece for Lemmas 5.6, 5.7, 5.8, and in turn the upper duality bound from Proposition 5.1.

LEMMA D.2. *Suppose Assumptions 1, 2 and 3 are satisfied. Then we have the limit*

$$\hat{\mathbb{Q}}[\tau^{\lambda,\mathrm{dual}} < \tau^{\Delta\varphi}] \longrightarrow 0 \quad \text{as } \lambda \to 0,$$

where $\tau^{\lambda,\mathrm{dual}}$ is defined in (5.16) and $\tau^{\Delta\varphi}$ is defined in (3.10).

PROOF. This is a simple consequence of Lemma D.1, and the inclusion

$$\begin{aligned}
&\{\tau^{\lambda,\mathrm{dual}} < \tau^{\Delta\varphi}\} \\
&\quad\subseteq \{(|B^{\lambda}|)^*_{\tau^{\Delta\varphi}} > \lambda^{-\kappa_5}\} \cup \left\{\left(\left|\frac{\mu^{\overline{S}^{\lambda},\hat{\mathbb{Q}}}\sqrt{c^S}}{c^{\overline{S}^{\lambda},S}}\right|\right)^*_{\tau^{\lambda,1}} > \lambda^{\kappa_6}\right\} \\
&\quad\cup \left\{\left(\left|\mathcal{E}\left(\int_0^{\cdot} \frac{\mu_t^{\overline{S}^{\lambda},\hat{\mathbb{Q}}}\sqrt{c_t^S}}{c_t^{\overline{S}^{\lambda},S}}\,dW^{S,\hat{\mathbb{Q}}}\right) - 1\right|\right)^*_{\tau^{\lambda,2}} > \lambda^{\kappa_6}\right\} \\
&\quad\cup \{(\lambda^{-\eta}|\overline{\Delta S}^{\lambda}|)^*_{\tau^{\lambda,1}} > \lambda^{\kappa_7}\} \cup \left\{\left(\frac{1}{c^S}|c^{\overline{\Delta S}^{\lambda}} + 2c^{S,\overline{\Delta S}^{\lambda}}|\right)^*_{\tau^{\lambda,1}} > \lambda^{\kappa_8}\right\} \\
&\quad\cup \{(\Lambda^{\frac{1}{p-1}})^*_T > \lambda^{-\kappa_9}\} \cup \left\{\left(\left|\frac{\mu^{\overline{S}^{\lambda},\hat{\mathbb{Q}}}\sqrt{c^S}}{c^{\overline{S}^{\lambda},S}} + \gamma\sqrt{c^S}\Delta\varphi^{\lambda}\right|\right)^*_{\tau^{\lambda,1}} > \lambda^{\kappa_{10}}\right\}.
\end{aligned}$$

□

Acknowledgments. The authors are grateful to Paolo Guasoni, Marko Weber and, in particular, to Ibrahim Ekren for fruitful discussions and valuable comments. They also thank two anonymous referees for their pertinent remarks.

Parts of this paper were written while Johannes Muhle-Karbe was visiting ETH Zürich; he is grateful to the Forschungsinstitut für Mathematik and H. Mete Soner for their hospitality.

The first author was supported in part by the Swiss National Science Foundation (SNF) under grant 175133.

The second author was supported in part by the Swiss National Science Foundation (SNF) under grant 150101.

REFERENCES

[1] AHRENS, L. (2015). On using shadow prices for the asymptotic analysis of portfolio optimization under proportional transaction costs. Ph.D. thesis, Christan-Albrechts-Universität zu Kiel.

[2] AHRENS, L. and KALLSEN, J. (2015). Portfolio optimization under small transaction costs: A convex duality approach. Preprint.

[3] ALMGREN, R. F. (2003). Optimal execution with nonlinear impact functions and trading-enhanced risk. *Appl. Math. Finance* **10** 1–18.

[4] ALMGREN, R. F. and CHRISS, N. (2001). Optimal execution of portfolio transactions. *J. Risk* **3** 5–40.

[5] ALMGREN, R. F. and LI, T. M. (2016). Option hedging with smooth market impact. *Market Microstucture Liq.* **2**.

[6] ALMGREN, R. F., THUM, C., HAUPTMANN, E. and LI, H. (2005). Direct estimation of equity market impact. *RISK* July.

[7] ALTAROVICI, A., MUHLE-KARBE, J. and SONER, H. M. (2015). Asymptotics for fixed transaction costs. *Finance Stoch.* **19** 363–414. MR3320325 https://doi.org/10.1007/s00780-015-0261-3

[8] BANK, P., SONER, H. M. and VOSS, M. (2017). Hedging with temporary price impact. *Math. Financ. Econ.* **11** 215–239. MR3604450 https://doi.org/10.1007/s11579-016-0178-4

[9] BARBERIS, N. (2000). Investing for the long run when returns are predictable. *J. Finance* **55** 225–264.

[10] BAYRAKTAR, E., CAYÉ, T. and EKREN, I. (2019). Asymptotics for small nonlinear price impact: A PDE approach to the multidimensional case. Preprint.

[11] BICHUCH, M. (2014). Pricing a contingent claim liability with transaction costs using asymptotic analysis for optimal investment. *Finance Stoch.* **18** 651–694. MR3232019 https://doi.org/10.1007/s00780-014-0233-z

[12] CAI, J., ROSENBAUM, M. and TANKOV, P. (2017). Asymptotic lower bounds for optimal tracking: A linear programming approach. *Ann. Appl. Probab.* **27** 2455–2514. MR3693531 https://doi.org/10.1214/16-AAP1264

[13] CAI, J., ROSENBAUM, M. and TANKOV, P. (2017). Asymptotic optimal tracking: Feedback strategies. *Stochastics* **89** 943–966. MR3733414 https://doi.org/10.1080/17442508.2017.1285304

[14] CAYÉ, T. (2017). Trading with small nonlinear price impact: Optimal execution and rebalancing of active investments. Ph.D. thesis, Eidgenössische Technische Hochschule Zürich.

[15] CAYÉ, T., HERDEGEN, M. and MUHLE-KARBE, J. (2020). Scaling limits of processes with fast nonlinear mean reversion. *Stochastic Process. Appl.* **130** 1994–2031. MR4074691 https://doi.org/j.spa.2019.06.008

[16] CONSTANTINIDES, G. M. (1986). Capital market equilibrium with transaction costs. *J. Polit. Econ.* **94** 842–862.

[17] CVITANIĆ, J. and KARATZAS, I. (1996). Hedging and portfolio optimization under transaction costs: A martingale approach. *Math. Finance* **6** 133–165. MR1384221 https://doi.org/10.1111/j.1467-9965.1996.tb00075.x

[18] DAVIS, M. H. A. (1997). Option pricing in incomplete markets. In *Mathematics of Derivative Securities* (*Cambridge*, 1995). *Publ. Newton Inst.* **15** 216–226. Cambridge Univ. Press, Cambridge. MR1491376

[19] DE LATAILLADE, J., DEREMBLE, C., POTTERS, M. and BOUCHAUD, J.-P. (2012). Optimal trading with linear costs. *J. Investment Strat.* **1** 91–115.

[20] DELBAEN, F., GRANDITS, P., RHEINLÄNDER, T., SAMPERI, D., SCHWEIZER, M. and STRICKER, C. (2002). Exponential hedging and entropic penalties. *Math. Finance* **12** 99–123. MR1891730 https://doi.org/10.1111/1467-9965.02001

[21] DOLINSKY, Y. and SONER, H. M. (2013). Duality and convergence for binomial markets with friction. *Finance Stoch.* **17** 447–475. MR3066984 https://doi.org/10.1007/s00780-012-0192-1

[22] FEODORIA, M. R. (2016). Optimal investment and utility indifference pricing in the presence of small fixed transaction costs. Ph.D. thesis, Christian-Albrechts-Universität zu Kiel.

[23] FRITTELLI, M. (2000). The minimal entropy martingale measure and the valuation problem in incomplete markets. *Math. Finance* **10** 39–52. MR1743972 https://doi.org/10.1111/1467-9965.00079

[24] FUKASAWA, M. (2011). Asymptotically efficient discrete hedging. In *Stochastic Analysis with Financial Applications. Progress in Probability* **65** 331–346. Birkhäuser/Springer Basel AG, Basel. MR3050797 https://doi.org/10.1007/978-3-0348-0097-6_19

[25] FUKASAWA, M. (2014). Efficient discretization of stochastic integrals. *Finance Stoch.* **18** 175–208. MR3146491 https://doi.org/10.1007/s00780-013-0215-6

[26] GARLEANU, N. and PEDERSEN, L. H. (2013). Dynamic trading with predictable returns and transaction costs. *J. Finance* **68** 2309–2340.

[27] GÂRLEANU, N. and PEDERSEN, L. H. (2016). Dynamic portfolio choice with frictions. *J. Econom. Theory* **165** 487–516. MR3540368 https://doi.org/10.1016/j.jet.2016.06.001

[28] GOBET, E. and LANDON, N. (2014). Almost sure optimal hedging strategy. *Ann. Appl. Probab.* **24** 1652–1690. MR3211007 https://doi.org/10.1214/13-AAP959
[29] GUASONI, P. and MUHLE-KARBE, J. (2015). Long horizons, high risk aversion, and endogenous spreads. *Math. Finance* **25** 724–753. MR3402460 https://doi.org/10.1111/mafi.12046
[30] GUASONI, P. and RÁSONYI, M. (2015). Hedging, arbitrage and optimality with superlinear frictions. *Ann. Appl. Probab.* **25** 2066–2095. MR3349002 https://doi.org/10.1214/14-AAP1043
[31] GUASONI, P. and WEBER, M. (2017). Dynamic trading volume. *Math. Finance* **27** 313–349. MR3635291 https://doi.org/10.1111/mafi.12099
[32] GUASONI, P. and WEBER, M. (2018). Nonlinear price Impact and portfolio choice. Preprint.
[33] HENDERSON, V. (2002). Valuation of claims on nontraded assets using utility maximization. *Math. Finance* **12** 351–373. MR1926237 https://doi.org/10.1111/1467-9965.12405
[34] HERDEGEN, M. and MUHLE-KARBE, J. (2018). Stability of Radner equilibria with respect to small frictions. *Finance Stoch.* **22** 443–502. MR3778362 https://doi.org/10.1007/s00780-018-0354-x
[35] HERDEGEN, M. and MUHLE-KARBE, J. (2019). Sensitivity of optimal consumption streams. *Stochastic Process. Appl.* **129** 1964–1992. MR3958420 https://doi.org/10.1016/j.spa.2018.06.011
[36] HODGES, S. and NEUBERGER, A. (1989). Optimal replication of contingent claims under transaction costs. *Rev. Futures Mark.* **8** 222–239.
[37] JANEČEK, K. and SHREVE, S. E. (2004). Asymptotic analysis for optimal investment and consumption with transaction costs. *Finance Stoch.* **8** 181–206. MR2048827 https://doi.org/10.1007/s00780-003-0113-4
[38] JOUINI, E. and KALLAL, H. (1995). Martingales and arbitrage in securities markets with transaction costs. *J. Econom. Theory* **66** 178–197. MR1338025 https://doi.org/10.1006/jeth.1995.1037
[39] KABANOV, Y. M. and STRICKER, C. (2002). On the optimal portfolio for the exponential utility maximization: Remarks to the six-author paper "Exponential hedging and entropic penalties" [Math. Finance **12** (2002), no. 2, 99–123; MR1891730 (2003b:91046)] by F. Delbaen, P. Grandits, T. Rheinländer, D. Samperi, M. Schweizer and C. Stricker. *Math. Finance* **12** 125–134. MR1891731 https://doi.org/10.1111/1467-9965.t01-1-02002
[40] KALLSEN, J. (2002). Derivative pricing based on local utility maximization. *Finance Stoch.* **6** 115–140. MR1885586 https://doi.org/10.1007/s780-002-8403-x
[41] KALLSEN, J. and MUHLE-KARBE, J. (2010). On using shadow prices in portfolio optimization with transaction costs. *Ann. Appl. Probab.* **20** 1341–1358. MR2676941 https://doi.org/10.1214/09-AAP648
[42] KALLSEN, J. and MUHLE-KARBE, J. (2015). Option pricing and hedging with small transaction costs. *Math. Finance* **25** 702–723. MR3402459 https://doi.org/10.1111/mafi.12035
[43] KALLSEN, J. and MUHLE-KARBE, J. (2017). The general structure of optimal investment and consumption with small transaction costs. *Math. Finance* **27** 659–703. MR3668154 https://doi.org/10.1111/mafi.12106
[44] KIM, T. and OMBERG, E. (1996). Dynamic nonmyopic portfolio behavior. *Rev. Financ. Stud.* **9** 141–161.
[45] KRAMKOV, D. and SÎRBU, M. (2006). Sensitivity analysis of utility-based prices and risk-tolerance wealth processes. *Ann. Appl. Probab.* **16** 2140–2194. MR2288717 https://doi.org/10.1214/105051606000000529
[46] LARSEN, K., MOSTOVYI, O. and ŽITKOVIĆ, G. (2018). An expansion in the model space in the context of utility maximization. *Finance Stoch.* **22** 297–326. MR3778357 https://doi.org/10.1007/s00780-017-0353-3
[47] LILLO, F., FARMER, J. D. and MANTEGNA, R. N. (2003). Master curve for price-impact function. *Nature* **421** 129–130.
[48] LIPTSER, R. S. and SHIRYAYEV, A. N. (2013). *Statistics of Random Processes. I. General Theory*, 2nd ed. *Applications of Mathematics* **5**. Springer, Berlin. MR0474486
[49] MARTIN, R. (2014). Optimal trading under proportional transaction costs. *RISK* August 54–59.
[50] MOREAU, L., MUHLE-KARBE, J. and SONER, H. M. (2017). Trading with small price impact. *Math. Finance* **27** 350–400. MR3635292 https://doi.org/10.1111/mafi.12098
[51] MOSTOVYI, O. and SÎRBU, M. (2019). Sensitivity analysis of the utility maximisation problem with respect to model perturbations. *Finance Stoch.* **23** 595–640. MR3968279 https://doi.org/10.1007/s00780-019-00388-1
[52] PESKIR, G. (2001). Bounding the maximal height of a diffusion by the time elapsed. *J. Theoret. Probab.* **14** 845–855. MR1860525 https://doi.org/10.1023/A:1017505509361
[53] PROTTER, P. E. (2004). *Stochastic Integration and Differential Equations: Stochastic Modelling and Applied Probability*, 2nd ed. *Applications of Mathematics (New York)* **21**. Springer, Berlin. MR2020294
[54] RHEINLÄNDER, T. and STEIGER, G. (2006). The minimal entropy martingale measure for general Barndorff–Nielsen/Shephard models. *Ann. Appl. Probab.* **16** 1319–1351. MR2260065 https://doi.org/10.1214/105051606000000240

[55] ROGERS, L. C. G. (2004). Why is the effect of proportional transaction costs $O(\delta^{2/3})$? In *Mathematics of Finance*. *Contemp*. *Math*. **351** 303–308. Amer. Math. Soc., Providence, RI. MR2076549 https://doi.org/10.1090/conm/351/06411

[56] ROSENBAUM, M. and TANKOV, P. (2014). Asymptotically optimal discretization of hedging strategies with jumps. *Ann*. *Appl*. *Probab*. **24** 1002–1048. MR3199979 https://doi.org/10.1214/13-AAP940

[57] SCHACHERMAYER, W. (2001). Optimal investment in incomplete markets when wealth may become negative. *Ann*. *Appl*. *Probab*. **11** 694–734. MR1865021 https://doi.org/10.1214/aoap/1015345346

[58] SCHACHERMAYER, W. (2003). A super-martingale property of the optimal portfolio process. *Finance Stoch*. **7** 433–456. MR2014244 https://doi.org/10.1007/s007800200096

[59] SCHACHERMAYER, W. and TEICHMANN, J. (2008). How close are the option pricing formulas of Bachelier and Black–Merton–Scholes? *Math*. *Finance* **18** 155–170. MR2380944 https://doi.org/10.1111/j.1467-9965.2007.00326.x

[60] SONER, H. M. and TOUZI, N. (2013). Homogenization and asymptotics for small transaction costs. *SIAM J*. *Control Optim*. **51** 2893–2921. MR3077898 https://doi.org/10.1137/120870165

[61] WHALLEY, A. E. and WILMOTT, P. (1997). An asymptotic analysis of an optimal hedging model for option pricing with transaction costs. *Math*. *Finance* **7** 307–324. MR1459062 https://doi.org/10.1111/1467-9965.00034